(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 July 2003 (17.07.2003)

PCT

(10) International Publication Number
WO 03/058871 A1

(51) International Patent Classification[7]: H04L 1/06, 5/02

(21) International Application Number: PCT/US02/41756

(22) International Filing Date: 31 December 2002 (31.12.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/042,529 8 January 2002 (08.01.2002) US

(71) Applicant: QUALCOMM INCORPORATED [US/US]; 5775 Morehouse Drive, San Diego, CA 92121 (US).

(72) Inventors: WALTON, Jay Rod; 7 Ledgewood Drive, Westford, MA 01886 (US). KETCHUM, John W.; 37 Candleberry Lane, Harvard, MA 01451 (US). WALLACE, Mark; 4 Madel Lane, Bedford, MA 01730 (US). MEDVEDEV, Irina; 68 Liberty Avenue, #2, Somerville, MA 02114 (US).

(74) Agents: WADSWORTH, Philip R. et al.; QUALCOMM Incorporated, 5775 Morehouse Drive, San Diego, CA 92121 (US).

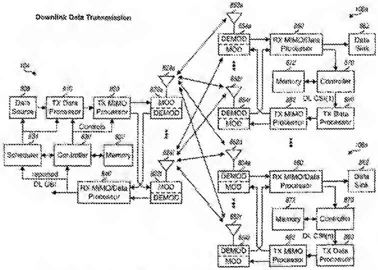
(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— with international search report

[Continued on next page]

(54) Title: RESSOURCE ALLOCATION FOR MIMO-OFDM COMMUNICATION SYSTEMS



(57) Abstract: Techniques to schedule terminals for data transmission on the downlink and/or uplink in a MIMO-OFDM system based on the spatial and/or frequency "signatures" of the terminals. A scheduler forms one or more sets of terminals for possible (downlink or uplink) data transmission for each of a number of frequency bands. One or more sub-hypotheses may further be formed for each hypothesis, with each sub-hypothesis corresponding to (1) specific assignments of transmit antennas to the terminal(s) in the hypothesis (for the downlinks) or (2) a specific order for processing the uplink data transmissions from the terminal(s) (for the uplink). The performance of each sub-hypothesis is then evaluated (e.g., based on one or more performance metrics). One sub-hypothesis is then selected for each frequency band based on the evaluated performance, and the one or more terminals in each selected sub-hypothesis are then scheduled for data transmission on the corresponding frequency band.

WO 03/058871 A1

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

RESOURCE ALLOCATION FOR MIMO-OFDM COMMUNICATION SYSTEMS

BACKGROUND

Field

[1001] The present invention relates generally to data communication, and more specifically to techniques for allocating resources in multiple-input multiple-output communication systems that utilize orthogonal frequency division multiplexing (i.e., MIMO-OFDM systems).

Background

[1002] A multiple-input multiple-output (MIMO) communication system employs multiple (N_T) transmit antennas and multiple (N_R) receive antennas for transmission of multiple independent data streams. In one MIMO system implementation, at any given moment, all of the data streams are used for a communication between a multiple-antenna base station and a single multiple-antenna terminal. However, in a multiple access communication system, the base station may also concurrently communicate with a number of terminals. In this case, each of the terminals employs a sufficient number of antennas such that it can transmit and/or receive one or more data streams.

[1003] The RF channel between the multiple-antenna array at the base station and the multiple-antenna array at a given terminal is referred to as a MIMO channel. The MIMO channel formed by the N_T transmit and N_R receive antennas may be decomposed into N_S independent channels, with $N_S \leq \min \{N_T, N_R\}$. Each of the N_S independent channels is also referred to as a spatial subchannel of the MIMO channel and corresponds to a dimension. The MIMO system can provide improved performance (e.g., increased transmission capacity) if the additional dimensionalities created by the multiple transmit and receive antennas are utilized.

[1004] A wideband MIMO system typically experiences frequency selective fading, which is characterized by different amounts of attenuation across the system bandwidth. This frequency selective fading causes inter-symbol interference (ISI), which is a phenomenon whereby each symbol in a received signal acts as distortion to subsequent

symbols in the received signal. This distortion degrades performance by impacting the ability to correctly detect the received symbols.

[1005] Orthogonal frequency division multiplexing (OFDM) may be used to combat ISI and/or for some other purposes. An OFDM system effectively partitions the overall system bandwidth into a number of (N_F) frequency subchannels, which may be referred to as subbands or frequency bins. Each frequency subchannel is associated with a respective subcarrier on which data may be modulated. The frequency subchannels of the OFDM system may also experience frequency selective fading, depending on the characteristics (e.g., the multipath profile) of the propagation path between the transmit and receive antennas. With OFDM, the ISI due to frequency selective fading may be combated by repeating a portion of each OFDM symbol (i.e., appending a cyclic prefix to each OFDM symbol), as is known in the art.

[1006] For a MIMO system that employs OFDM (i.e., a MIMO-OFDM system), N_F frequency subchannels are available for each of the N_S spatial subchannels of a MIMO channel. Each frequency subchannel of each spatial subchannel may be referred to as a transmission channel. Up to $N_F \cdot N_S$ transmission channels may be available for use at any given moment for communication between the multiple-antenna base station and the multiple-antenna terminal.

[1007] The MIMO channel between the base station and each terminal typically experiences different link characteristics and may thus be associated with different transmission capabilities. Moreover, each spatial subchannel may further experience frequency selective fading, in which case the frequency subchannels may also be associated with different transmission capabilities. Thus, the transmission channels available to each terminal may have different effective capacities. Efficient use of the available resources and higher throughput may be achieved if the $N_F \cdot N_S$ available transmission channels are effectively allocated such that these channels are utilized by a "proper" set of one or more terminals in the MIMO-OFDM system.

[1008] There is therefore a need in the art for techniques to allocate resources in a MIMO-OFDM system to provide high system performance.

SUMMARY

[1009] Techniques are provided herein to schedule terminals for data transmission on the downlink and/or uplink based on the spatial and/or frequency "signatures" of the terminals. In a MIMO-OFDM system, each "active" terminal desiring data transmission in an upcoming time interval may be associated with transmission channels having different capabilities due to different link conditions experienced by the terminal. Various scheduling schemes are provided herein to select a "proper" set of one or more terminals for data transmission on each frequency band and to assign the available transmission channels to the selected terminals such that system goals (e.g., high throughput, fairness, and so on) are achieved.

[1010] A scheduler may be designed to form one or more sets of terminals for possible (downlink or uplink) data transmission for each of a number of frequency bands. Each set includes one or more active terminals and corresponds to a hypothesis to be evaluated. Each frequency band corresponds to a group of one or more frequency subchannels in the MIMO-OFDM system. The scheduler may further form one or more sub-hypotheses for each hypothesis. For the downlink, each sub-hypothesis may correspond to specific assignments of a number of transmit antennas at the base station to the one or more terminals in the hypothesis. And for the uplink, each sub-hypothesis may correspond to a specific order for processing the uplink data transmissions from the one or more terminals in the hypothesis. The performance of each sub-hypothesis is then evaluated (e.g., based on one or more performance metrics, such as a performance metric indicative of the overall throughput for the terminals in the hypothesis). One sub-hypothesis is then selected for each frequency band based on the evaluated performance, and the one or more terminals in each selected sub-hypothesis are then scheduled for data transmission on the corresponding frequency band.

[1011] The set of one or more terminals scheduled for (downlink or uplink) data transmission on each frequency band may include multiple SIMO terminals, a single MIMO terminal, multiple MISO terminals, or a combination of SIMO, MISO, and MIMO terminals. A SIMO terminal is one scheduled for data transmission via a single spatial subchannel in the MIMO-OFDM system and which employs multiple receive antennas and a single transmit antenna, a MISO terminal is one utilizing a single receive antenna to receive a transmission utilizing a single spatial subchannel, and a MIMO terminal is one scheduled for data transmission via two or more spatial subchannels.

Each SIMO, MISO, or MIMO terminal may be assigned with one or multiple frequency bands for data transmission. The available transmission channels are assigned to the terminals such that the system goals are achieved.

[1012] Details of various aspects, embodiments, and features of the invention are described below. The invention further provides methods, computer products, schedulers, base stations, terminals, systems, and apparatuses that implement various aspects, embodiments, and features of the invention, as described in further detail below.

BRIEF DESCRIPTION OF THE DRAWINGS

[1013] The features, nature, and advantages of the present invention will become more apparent from the detailed description set forth below when taken in conjunction with the drawings in which like reference characters identify correspondingly throughout and wherein:

[1014] FIG. 1 is a diagram of a MIMO-OFDM system;

[1015] FIG. 2 is a flow diagram of a process to schedule terminals for downlink data transmission;

[1016] FIG. 3 is a flow diagram of a process to assign transmit antennas to terminals using a "max-max" criterion;

[1017] FIG. 4 is a flow diagram for a priority-based downlink scheduling scheme whereby a set of N_T highest priority terminals is considered for scheduling;

[1018] FIG. 5 is a flow diagram of a process to schedule terminals for uplink transmission;

[1019] FIGS. 6A and 6B are flow diagrams for two successive cancellation receiver processing schemes whereby the processing order is (1) imposed by an ordered set of terminals and (2) determined based on the post-processed SNRs, respectively;

[1020] FIG. 7 is a flow diagram for a priority-based uplink scheduling scheme whereby a set of N_T highest priority terminals is considered for scheduling;

[1021] FIGS. 8A and 8B are block diagrams of a base station and two terminals for downlink and uplink data transmission, respectively;

[1022] FIG. 9 is a block diagram of an embodiment of a transmitter unit; and

[1023] FIGS. 10A and 10B are block diagrams of two embodiments of a receiver unit without and with successive cancellation receiver processing, respectively;

DETAILED DESCRIPTION

[1024] FIG. 1 is a diagram of a multiple-input multiple-output communication system 100 that utilizes orthogonal frequency division multiplexing (i.e., a MIMO-OFDM system). MIMO-OFDM system 100 employs multiple (N_T) transmit antennas and multiple (N_R) receive antennas for data transmission. MIMO-OFDM system 100 may be a multiple-access communication system having one or more base stations (BS) 104 that can concurrently communicate with one or more terminals (T) 106 (only one base station is shown in FIG. 1 for simplicity). The base stations may also be referred to as access points, UTRAN, or some other terminology, and the terminals may also be referred to as handsets, mobile stations, remote stations, user equipment, or some other terminology.

[1025] Each base station 104 employs multiple antennas and represents the multiple-input (MI) for downlink transmissions from the base station to the terminals and the multiple-output (MO) for uplink transmissions from the terminals to the base station. A set of one or more "communicating" terminals 106 collectively represents the multiple-output for downlink transmissions and the multiple-input for uplink transmissions. As used herein, a communicating terminal is one that transmits and/or receives user-specific data to/from the base station, and an "active" terminal is one that desires downlink and/or uplink data transmission in an upcoming or future time slot. Active terminals may include terminals that are currently communicating.

[1026] For the example shown in FIG. 1, base station 104 concurrently communicates with terminals 106a through 106d (as indicated by the solid lines) via the multiple antennas available at the base station and the one or more antennas available at each communicating terminal. Terminals 106e through 106h may receive pilots and/or other signaling information from base station 104 (as indicated by the dashed lines), but are not transmitting or receiving user-specific data to/from the base station.

[1027] For the downlink, the base station employs N_T antennas and each communicating terminal employs 1 or N_R antennas for reception of one or more data streams from the base station. In general, N_R can be any integer two or greater. A MIMO channel formed by the N_T transmit antennas and N_R receive antennas may be decomposed into N_S independent channels, with $N_S \leq \min \{N_T, N_R\}$. Each such independent channel may be referred to as a spatial subchannel of the MIMO channel.

[1028] For the downlink, the number of receive antennas at a communicating terminal may be equal to or greater than the number of transmit antennas at the base station (i.e., $N_R \geq N_T$). For such a terminal, the number of spatial subchannels is limited by the number of transmit antennas at the base station. Each multi-antenna terminal communicates with the base station via a respective MIMO channel formed by the base station's N_T transmit antennas and its own N_R receive antennas. However, even if multiple multi-antenna terminals are selected for downlink data transmission, only N_S spatial subchannels are available regardless of the number of terminals receiving the downlink transmission. The terminals to be considered for downlink data transmission need not all be equipped with equal number of receive antennas.

[1029] For the downlink, the number of receive antennas at a communicating terminal may also be less than the number of transmit antennas at the base station (i.e., $N_R < N_T$). In particular, a MISO terminal employs a single receive antenna ($N_R = 1$) for downlink data transmission. The base station may then employ beam steering and space division multiple access (SDMA) to communicate simultaneously with a number of MISO terminals, as described below.

[1030] For the uplink, each communicating terminal may employ a single antenna or multiple antennas for uplink data transmission. Each terminal may also utilize all or only a subset of its available antennas for uplink transmission. At any given moment, the N_T transmit antennas for the uplink are formed by all antennas used by one or more communicating terminals. The MIMO channel is then formed by the N_T transmit antennas from all communicating terminals and the base station's N_R receive antennas. The number of spatial subchannels is limited by the number of transmit antennas, which in turn is limited by the number of receive antennas at the base station (i.e., $N_S \leq \min(N_T, N_R)$).

[1031] With SDMA, the "spatial signatures" associated with different terminals are exploited to allow multiple terminals to operate simultaneously on the same channel, which may be a time slot, a frequency band, a code channel, and so on. A spatial signature constitutes a complete RF characterization of the propagation path between each transmit-receive antenna pair to be used for data transmission. On the downlink, the spatial signatures may be derived at the terminals and reported to the base station. The base station may then process these spatial signatures to select terminals for data transmission on the same channel, and to derive mutually "orthogonal" steering vectors

for each of the independent data streams to be transmitted to the selected terminals. On the uplink, the base station may derive the spatial signatures of the different terminals. The base station may then process these signatures to schedule terminals for data transmission and to further process the transmissions from the scheduled terminals to separately demodulate each transmission

[1032] If the terminals are equipped with multiple receive antennas such that $N_R \geq N_T$, then the base station does not need the spatial signatures of the terminals in order to obtain the benefit of SDMA. All that may be needed at the base station is information from each terminal indicating the "post-processed" SNR associated with the signal from each base station transmit antenna, after demodulation at the terminal. The SNR estimation process may be facilitated by periodically transmitting a pilot from each base station transmit antenna, as described below.

[1033] As used herein, a SIMO terminal is one designated (or scheduled) to transmit and/or receive data via a single spatial subchannel and which employs multiple receive antennas for data transmission, a MISO terminal is one designated to receive a data transmission via a single spatial subchannel and which employs a single receive antenna, and a MIMO terminal is one designated to transmit and/or receive data via multiple spatial subchannels. For the downlink, a SIMO terminal may receive a data transmission from a single transmit antenna at the base station, and a MISO terminal may receive a data transmission via a beam formed by the N_T transmit antennas at the base station. And for the uplink, the SIMO terminal may transmit data from one antenna at the terminal.

[1034] For the MIMO-OFDM system, each spatial subchannel is further partitioned into N_F frequency subchannels. Each frequency subchannel of each spatial subchannel may be referred to as a transmission channel. For both the downlink and uplink, the N_T transmit antennas may thus be used to transmit up to $N_F{\cdot}N_S$ independent data streams on the $N_F{\cdot}N_S$ transmission channels. Each independent data stream is associated with a particular "rate", which is indicative of various transmission parameter values such as, for example, a specific data rate, a specific coding scheme, a specific modulation scheme, and so on, used for the data stream. The rate is typically determined by the capacity of the one or more transmission channels used to transmit the data stream.

Multi-User OFDM System

[1035] For a multiple-access OFDM system without MIMO capability, the overall system bandwidth, W, is divided into N_F orthogonal frequency subchannels, with each such subchannel having a bandwidth of W/N_F. For this system, a number of terminals may share the available spectrum via time division multiplexing (TDM). In a "pure" TDM scheme, a single terminal may be assigned the entire system bandwidth, W, for each fixed time interval, which may be referred to as a time slot. Terminals may be scheduled for data transmission by allocating time slots on a demand basis. Alternatively, for this OFDM system, it is possible to assign only a fraction, N_A, of the N_F frequency subchannels to a given terminal for a given time slot, thus making the remaining (N_F - N_A) frequency subchannels in the same time slot available to other terminals. In this way, the TDM access scheme is converted into a hybrid TDM/FDM access scheme.

[1036] Allocating different frequency subchannels to different terminals may provide improved performance for frequency selective channel. In the pure TDM scheme whereby all N_F frequency subchannels are allocated to a single terminal for a given time slot, it is possible that some of the frequency subchannels associated with this terminal could be faded, thereby resulting in low SNR and poor throughput for these faded subchannels. However, these same frequency subchannels may have high SNR for another terminal in the system since the RF channel is likely to be uncorrelated from terminal to terminal. If a scheduler has knowledge of the SNR for each active terminal and for all N_F frequency subchannels, then it may be possible to maximize system throughput by allocating each of the N_F frequency subchannels to the terminal achieving the best SNR for that subchannel. In practice, certain minimum performance requirements typically need to be met for all terminals so that the scheduler would need to observe some fairness criteria to ensure that the terminals in the best locations do not continually "hog" the resources.

[1037] The pure TDM scheduling scheme described above can assign time slots to terminals that have favorable fading conditions. For improved performance, the scheduler can further consider allocating frequency subchannels to terminals in each time slot and possibly allocating transmit power per subchannel. The ability to allocate transmit power provides an additional degree of scheduling flexibility that may be used to improve performance (e.g., to increase throughput).

Single-User MIMO-OFDM System

[1038] For the MIMO-OFDM system, the N_F frequency subchannels may be used to transmit up to N_F independent data streams on each of the N_S spatial subchannels. The total number of transmission channels is thus $N_C = N_F \cdot N_S$. For the pure TDM scheme, the N_C transmission channels may be allocated to a single terminal for each time slot.

[1039] The N_C transmission channels may be associated with different SNRs and may have different transmission capabilities. A fraction of the transmission channels may achieve poor SNR. In one scheme, additional redundancy (e.g., a lower rate code) may be used for the transmission channels with poor SNR to achieve the target packet error rate (PER). The additional redundancy effectively reduces throughput. In another scheme, some or all of the transmission channels with poor SNR may be eliminated from use, and only a subset of the available frequency subchannels is selected for use for each spatial subchannel.

[1040] The total available transmit power may be allocated uniformly or non-uniformly across the transmission channels to improve throughput. For example, the total available transmit power for each transmit antenna may be allocated in a uniform or non-uniform manner across the frequency subchannels selected for use for that transmit antenna. In this way, transmit power is not wasted on transmission channels that provide little or no information to allow the receiver to recover the transmitted data. The frequency subchannel selection and the power allocation may be implemented on a per-transmit antenna basis whereby (1) all or a subset of the N_F frequency subchannels for each transmit antenna may be selected for use, and (2) the transmit power available for each transmit antenna may be uniformly or non-uniformly allocated across the selected frequency subchannels.

[1041] The technique used to process the received signals at the receiver can have an impact on which transmission channels get selected for use. If a successive equalization and interference cancellation (or "successive cancellation") receiver processing technique (described below) is used at the receiver, then it may be advantageous to disable certain transmit antennas in order to increase throughput on the link. In this case, the receiver can determine which subset of transmit antennas should be used for data transmission and can provide this information to the transmitter via a feedback channel. If the RF channel experiences frequency selective fading, then the

set of transmit antennas used for one frequency subchannel may not be the best set to use for another frequency subchannel. In this case, the scheduler can select a proper set of transmit antennas to use on a per frequency subchannel basis in order to improve throughput.

Multi-User MIMO-OFDM System

[1042] Various techniques are described above for (1) allocating different frequency subchannels to different terminals in a multi-user OFDM system, and (2) allocating transmission channels to a single terminal in a single-user MIMO-OFDM system. These techniques may also be used to allocate resources (e.g., transmission channels and transmit power) to multiple terminals in a multiple-access MIMO-OFDM system. Various scheduling schemes may be designed to achieve high system throughput by utilizing these and possibly other techniques for the multi-user environment.

[1043] The system resources may be allocated by selecting the "best" set of terminals for data transmission such that high throughput and/or some other criteria are achieved. With frequency selective fading, the resource allocation may be performed for each group of one or more frequency subchannels. Resource allocation for each fractional portion of the overall system bandwidth may provide additional gains over a scheme that attempts to maximize throughput on the total system bandwidth basis (i.e., as would be the case for a single carrier MIMO system).

[1044] If the entire system bandwidth is treated as a single frequency channel (e.g., as in a single carrier MIMO system), then the maximum number of terminals that may be scheduled to transmit simultaneously is equal to the number of spatial subchannels, which is $N_S \leq \min\{N_R, N_T\}$. If the system bandwidth is divided into N_F frequency channels (e.g., as in a MIMO-OFDM system), then the maximum number of terminals that may be scheduled to transmit simultaneously is $N_F \cdot N_S$, since each transmission channel (i.e., each frequency subchannel of each spatial subchannel) may be allocated to a different terminal. And if the system bandwidth is divided into N_G groups of frequency subchannels, then the maximum number of terminals that may be scheduled to transmit simultaneously is $N_G \cdot N_S$, since each frequency subchannel group of each spatial subchannel may be allocated to a different terminal. If the number of terminals is less than the maximum number permitted, then multiple transmission channels may be allocated to a given terminal.

[1045] Various operating modes may be supported by the MIMO-OFDM system. In a MIMO mode, all spatial subchannels of a particular frequency subchannel group are allocated to a single MIMO terminal. Multiple MIMO terminals may still be supported concurrently via the N_G frequency subchannel groups. In an N-SIMO mode, the N_S spatial subchannels of a particular frequency subchannel group are allocated to a number of distinct SIMO terminals, with each SIMO terminal being assigned one spatial subchannel. A given SIMO terminal may be assigned one or more frequency subchannel groups of a particular spatial subchannel. In an N-MISO mode (which may also be referred to a multi-user beam-steering mode), the N_S spatial subchannels of a particular frequency subchannel group are allocated to a number of distinct MISO terminals, with each MISO terminal being assigned one spatial subchannel. Full characterization of the transmit-receive antenna paths may be used to derive distinctive beams for the data transmission to these MISO terminals. Similarly, a given MISO terminal may be assigned one or more frequency subchannel groups of a particular spatial subchannel. And in a mixed mode, the N_S spatial subchannels for a particular frequency subchannel group may be allocated to a combination of SIMO, MISO, and MIMO terminals, with multiple types of terminals being concurrently supported. Any combination of operating modes may be supported for a particular time slot. For example, the MIMO mode may be supported for the first frequency subchannel group, the N-SIMO mode may be supported for the second frequency subchannel group, the N-MISO mode may be supported for the third frequency subchannel group, the mixed mode may be supported for the fourth frequency subchannel group, and so on. By communicating simultaneously with multiple SIMO terminals, multiple MISO terminals, one or more MIMO terminals, or a combination of SIMO, MISO, and MIMO terminals, the system throughput may be increased.

[1046] If the propagation environment has sufficient scattering, then MIMO receiver processing techniques may be used to efficiently exploit the spatial dimensionalities of the MIMO channel to increase transmission capacity. MIMO receiver processing techniques may be used whether the base station is communicating with one or multiple terminals simultaneously. For the downlink, from a terminal's perspective, the same receiver processing techniques may be used to process N_T different signals intended for that terminal (if it is a MIMO terminal) or just one of the N_T signals (if it is a SIMO terminal). If successive cancellation receiver processing is to

be used at the terminals, then certain restrictions may apply since a data stream assigned to one terminal may not be detected error-free by another terminal. And for the uplink, from the base station's perspective, there is no discernable difference in processing N_T different signals from a single MIMO terminal versus processing one signal from each of N_T different SIMO terminals.

[1047] As shown in FIG. 1, the terminals may be randomly distributed in the base station's coverage area (or "cell") or may be co-located. For a wireless communication system, the link characteristics typically vary over time due to a number of factors such as fading and multipath. At a particular instant in time, the response for a MIMO channel formed by an array of N_T transmit antennas and an array of N_R receive antennas may be characterized by a matrix $\underline{\mathbf{H}}(k)$ whose elements are composed of independent Gaussian random variables, as follows:

$$\underline{\mathbf{H}}(k)=\begin{bmatrix} h_{1,1}(k) & h_{1,2}(k) & \Lambda & h_{1,N_T}(k) \\ h_{2,1}(k) & h_{2,2}(k) & \Lambda & h_{2,N_T}(k) \\ M & M & & M \\ h_{N_R,1}(k) & h_{N_R,2}(k) & \Lambda & h_{N_R,N_T}(k) \end{bmatrix} \qquad \text{Eq (1)}$$

For the downlink, the array of N_T transmit antennas is at the base station, and the array of N_R receive antennas may be formed at a single SIMO or MIMO terminal (for the N-SIMO or MIMO mode) or at multiple MISO terminals (for the N-MISO mode). And for the uplink, the transmit antenna array is formed by the antennas used by all communicating terminals, and the receive antenna array is at the base station. In equation (1), $\underline{\mathbf{H}}(k)$ is the channel response matrix for the MIMO channel for the k-th frequency subchannel group, and $h_{i,j}(k)$ is the coupling (i.e., the complex gain) between the j-th transmit antenna and the i-th receive antenna for the k-th frequency subchannel group.

[1048] Each frequency subchannel group may include one or more frequency subchannels and corresponds to a particular frequency band of the overall system bandwidth. Depending on the particular system design, there may be (1) only one group with all N_F frequency subchannels, or (2) N_F groups, with each group having a single frequency subchannel, or (3) any number of groups between 1 and N_F. The number of frequency subchannel groups, N_G, can thus range between 1 and N_F, inclusive (i.e., $1 \leq N_G \leq N_F$). Each group may include any number of frequency subchannels, and the N_G

groups may include the same or different number of frequency subchannels. Moreover, each group may include any combination of frequency subchannels (e.g., the frequency subchannels for a group need not be adjacent to one another).

[1049] As shown in equation (1), the MIMO channel response for each frequency subchannel group may be represented with a respective matrix $\underline{\mathbf{H}}(k)$ having $N_R \times N_T$ elements corresponding to the number of receive antennas and the number of transmit antennas. Each element of the matrix $\underline{\mathbf{H}}(k)$ describes the response for a respective transmit-receive antenna pair for the k-th frequency subchannel group. For a flat fading channel (or when $N_G = 1$), one complex value may be used for the entire system bandwidth (i.e., for all N_F frequency subchannels) for each transmit-receive antenna pair.

[1050] In an actual operating environment, the channel response typically varies across the system bandwidth, and a more detailed channel characterization may be used for the MIMO channel. Thus, for a frequency selective fading channel, one channel response matrix $\underline{\mathbf{H}}(k)$ may be provided for each frequency subchannel group. Alternatively, a channel impulse response matrix, $\underline{\hat{\mathcal{H}}}(n)$, may be provided for the MIMO channel, with each element of this matrix corresponding to a sequence of values indicative of the sampled impulse response for a respective transmit-receive antenna pair.

[1051] The receiver may periodically estimate the channel response for each transmit-receive antenna pair. The channel estimates may be facilitated in a number of ways such as, for example, with the use of pilot and/or data decision directed techniques known in the art. The channel estimates may comprise the complex-value channel response estimate (e.g., the gain and phase) for each frequency subchannel group of each transmit-receive antenna pair, as shown in equation (1). The channel estimates provide information on the transmission characteristics of (e.g., what data rate is supportable by) each spatial subchannel for each frequency subchannel group.

[1052] The information given by the channel estimates may also be distilled into (1) a post-processed signal-to-noise-and-interference ratio (SNR) estimate (described below) for each spatial subchannel of each frequency subchannel group, and/or (2) some other statistic that allows the transmitter to select the proper rate for each independent data stream. This process of deriving the essential statistic may reduce the amount of

data required to characterize a MIMO channel. The complex channel gains and the post-processed SNRs represent different forms of channel state information (CSI) that may be reported by the receiver to the transmitter. For time division duplexed (TDD) systems, the transmitter may be able to derive or infer some of the channel state information based on transmission (e.g., a pilot) from the receiver since there may be sufficient degree of correlation between the downlink and uplink for such systems, as described below. Other forms of CSI may also be derived and reported and are described below.

[1053] The aggregate CSI received from the receivers may be used to achieve high throughput by assigning a proper set of one or more terminals to the available transmission channels such that they are allowed to communicate simultaneously with the base station. A scheduler can evaluate which specific combination of terminals provides the best system performance (e.g., the highest throughput) subject to any system constraints and requirements.

[1054] By exploiting the spatial and frequency "signatures" of the individual terminals (i.e., their channel response estimates, which may be a function of frequency), the average throughput can be increased relative to that achieved by a single terminal. Furthermore, by exploiting multi-user diversity, the scheduler can identify combinations of "mutually compatible" terminals that can be allowed to communicate at the same time on the same channel, effectively enhancing system capacity relative to single-user scheduling and random scheduling for multiple users.

[1055] The terminals may be scheduled for data transmission based on various factors. One set of factors may relate to system constraints and requirements such as the desired quality of service (QoS), maximum latency, average throughput, and so on. Some or all of these factors may need to be satisfied on a per terminal basis (i.e., for each terminal) in a multiple-access communication system. Another set of factors may relate to system performance, which may be quantified by an average system throughput or some other indications of performance. These various factors are described in further detail below.

[1056] For the downlink, the scheduler may (1) select the "best" set of one or more terminals for data transmission, (2) assign the available transmission channels to the selected terminals, (3) allocate transmit power uniformly or non-uniformly across the assigned transmission channels, and (4) determine the proper rate for each independent

data stream to be transmitted to the selected terminals. For the uplink, the scheduler may (1) select the best set of one or more terminals for data transmission, (2) assign the available transmission channels to the selected terminals, (3) determine the proper order for processing the data streams from these selected terminals (if the successive cancellation receiver processing technique is used at the base station), and (4) determine the rate for each independent data stream from the selected terminals. Various details of the resource allocation for the downlink and uplink are described below.

[1057] To simplify the scheduling, the terminals may be allocated transmission channels (and possibly transmit power) based on their priority. Initially, the active terminals may be ranked by their priority, which may be determined based on various factors, as described below. The N_X highest priority terminals may then be considered in each scheduling interval. This then allows the scheduler to allocate the available transmission channels to just N_X terminals instead of all active terminals. The resource allocation may be further simplified by (1) selecting $N_X = N_S$ and assigning each terminal with all frequency subchannels of one spatial subchannel, or (2) selecting $N_X = N_G$ and assigning each terminal with all spatial subchannels of one frequency subchannel group, or (3) making some other simplification. The gains in throughput even with some of these simplifications may be substantial compared to the pure TDM scheduling scheme that allocates all transmission channels to a single terminal for each time slot, particularly if independent frequency selective fading of the N_X terminals is considered in the resource allocation.

[1058] For simplicity, several assumptions are made in the following description. First, it is assumed that the average received power for each independent data stream may be adjusted to achieve a particular target energy-per-bit-to-total-noise-and-interference ratio (E_b/N_t) after signal processing at the receiver (which is the terminal for a downlink transmission and the base station for an uplink transmission). This target E_b/N_t is often referred to as a power control setpoint (or simply, the setpoint) and is selected to provide a particular level of performance (e.g., a particular PER). The setpoint may be achieved by a closed-loop power control mechanism that adjusts the transmit power for each data stream (e.g., based on a power control signal from the receiver). For simplicity, a common setpoint may be used for all data streams received at the receiver. Alternatively, a different setpoint may be used for each data stream, and the techniques described herein may be generalized to cover this operating mode. Also,

for the uplink, it is assumed that simultaneous transmissions from different terminals are synchronized so that the transmissions arrive at the base station within a prescribed time window.

[1059] For simplicity, the number of receive antennas is assumed to be equal to the number of transmit antennas (i.e., $N_R = N_T$) for the following description of the N-SIMO and MIMO modes. This is not a necessary condition since the analysis applies for the case where $N_R \geq N_T$. For the N-MISO mode, the number of receive antennas at each MISO terminal is assumed to be equal to one (i.e., $N_R = 1$). Also for simplicity, the number of spatial subchannels is assumed to be equal to the number of transmit antennas (i.e., $N_S = N_T$).

Downlink Resource Allocation

[1060] Resource allocation for the downlink comprises (1) selecting one or more sets of terminals for evaluation, (2) assigning the available transmission channels to the terminals in each set and evaluating performance, and (3) identifying the best set of terminals and their transmission channel assignments. Each set may include a number of SIMO terminals, a number of MISO terminals, one or more MIMO terminals, or a combination of SIMO, MISO, and MIMO terminals. All or only a subset of the active terminals may be considered for evaluation, and these terminals may be selected to form one or more sets to be evaluated. Each terminal set corresponds to a hypothesis. For each hypothesis, the available transmission channels may be assigned to the terminals in the hypothesis based on any one of a number of channel assignment schemes. The terminals in the best hypothesis may then be scheduled for data transmission in an upcoming time slot. The flexibility in both selecting the best set of terminals for data transmission and assigning the transmission channels to the selected terminals allows the scheduler to exploit multi-user diversity environment to achieve high performance in both flat fading and frequency selective fading channels.

[1061] In order to determine the "optimum" transmission to a set of terminals, SNRs or some other sufficient statistics may be provided for each terminal. For the N-SIMO and MIMO modes, where ($N_R \geq N_T$), the spatial processing may be performed by at the SIMO and MIMO terminals to separate out the transmitted signals, and the base station does not need the spatial signatures of the terminals in order to simultaneously transmit multiple data streams on the available spatial subchannels. All that may be

needed at the base station is the post-processed SNR associated with the signal from each base station transmit antenna. For clarity, downlink scheduling for SIMO and MIMO terminals is described first, and downlink scheduling for MISO terminals is described subsequently.

Downlink Scheduling for SIMO and MIMO Terminals

[1062] The scheduling for SIMO and MIMO terminals may be performed based on various types of channel state information, including full-CSI (e.g., complex channel gains) and partial-CSI (e.g., SNRs). If the statistic to be used for scheduling terminals is SNR, then for each set of one or more terminals to be evaluated for data transmission in an upcoming time slot, a hypothesis matrix $\underline{\Gamma}(k)$ of post-processed SNRs for this terminal set for the k-th frequency subchannel group may be expressed as:

$$\underline{\Gamma}(k) = \begin{bmatrix} \gamma_{1,1}(k) & \gamma_{2,1}(k) & \cdots & \gamma_{N_T,1}(k) \\ \gamma_{1,2}(k) & \gamma_{2,2}(k) & \cdots & \gamma_{N_T,2}(k) \\ \vdots & \vdots & & \vdots \\ \gamma_{1,N_T}(k) & \gamma_{2,N_T}(k) & \cdots & \gamma_{N_T,N_T}(k) \end{bmatrix}, \qquad \text{Eq (2)}$$

where $\gamma_{i,j}(k)$ is the post-processed SNR for a data stream (hypothetically) transmitted from the j-th transmit antenna to the i-th terminal for the k-th frequency subchannel group. A set of N_G such matrices $\underline{\Gamma}(k)$, for $1 \leq k \leq N_G$, would then characterize the entire frequency and spatial dimensions for this set of terminals.

[1063] At each terminal in the set being evaluated, N_T data streams may be (hypothetically) transmitted from the N_T transmit antennas for each frequency subchannel group and received by that terminal's N_R receive antennas. The N_R received signals at the terminal may be processed using spatial or space-time equalization to separate out the N_T transmitted data streams for each frequency subchannel group, as described below. The SNR of a post-processed data stream (i.e., after the equalization) may be estimated and comprises the post-processed SNR for that data stream. For each terminal, a set of N_T post-processed SNRs may be provided for the N_T data streams that may be received by that terminal for each of the N_G frequency subchannel groups.

[1064] In the N-SIMO mode, the N_T rows of the hypothesis matrix $\underline{\Gamma}(k)$ correspond to N_T vectors of SNRs for N_T different terminals for the k-th frequency subchannel

group. In this mode, each row of the hypothesis matrix $\underline{\Gamma}(k)$ gives the SNR of each of the N_T (hypothetical) data streams from the N_T transmit antennas for the k-th frequency subchannel group for one SIMO terminal. In the MIMO mode, the N_T rows of the hypothesis matrix $\underline{\Gamma}(k)$ correspond to a single vector of SNRs for a single MIMO terminal. This SNR vector includes the SNRs for the N_T data streams for the k-th frequency subchannel group, and may be replicated N_T times to form the matrix $\underline{\Gamma}(k)$. And in the mixed mode, for a particular MIMO terminal to be potentially assigned with two or more spatial subchannels for the k-th frequency subchannel group, that terminal's vector of SNRs may be replicated such that the SNR vector appears in as many rows of the hypothesis matrix $\underline{\Gamma}(k)$ as the number of spatial subchannels to be assigned to the terminal (i.e., one row per spatial subchannel).

[1065] Alternatively, for all operating modes, one row in the hypothesis matrix $\underline{\Gamma}(k)$ may be used for each SIMO or MIMO terminal, and the scheduler may be designed to mark and evaluate these different types of terminals accordingly. For the following description, the hypothesis matrix $\underline{\Gamma}(k)$ is assumed to include SNR vectors for N_T terminals, where an SIMO terminal is represented as a single terminal in the matrix and a MIMO terminal may be represented as two or more of the N_T terminals in the matrix.

[1066] If the successive cancellation receiver processing technique is used at a terminal to process the received signals, then the post-processed SNR achieved at the terminal for each transmitted data stream for a particular frequency subchannel group depends on the order in which the transmitted data streams are detected (i.e., demodulated and decoded) to recover the transmitted data, as described below. In this case, a number of sets of SNRs may be provided for each terminal for a number of possible detection orderings. Multiple hypothesis matrices $\underline{\Gamma}(k)$ may then be formed for each frequency subchannel group of each set of terminals, and these matrices may be evaluated to determine which specific combination of terminals and detection ordering provides the best system performance.

[1067] In any case, each hypothesis matrix $\underline{\Gamma}(k)$ includes the post-processed SNRs for a given frequency subchannel group for a specific set of terminals (i.e., hypothesis) to be evaluated. These post-processed SNRs represent the SNRs achievable by the terminals and are used to evaluate the hypothesis.

[1068] For the N-SIMO and MIMO modes, each transmit antenna in the base station's antenna array may be used to transmit a different data stream on each frequency subchannel group using channel state information (e.g., SNRs or some other statistics) derived by the terminals in the coverage area. High performance is achieved on the basis of the CSI, which is used to schedule terminals and process data.

[1069] Various downlink scheduling schemes may be used to allocate resources (e.g., transmission channels) to the active terminals. These various schemes include (1) an "exhaustive" scheduling scheme that can assign each transmission channel to a specific terminal such that "optimum" performance, as determined by some metrics, is achieved, (2) a priority-based scheduling scheme that assigns transmission channels based on the priority of the active terminals, (3) a FDM-TDM scheduling scheme that assigns all spatial subchannels of each frequency subchannel group to a specific terminal, and (4) a SDMA-TDM scheduling scheme that assigns all frequency subchannels of each spatial subchannel to a specific terminal. These various downlink scheduling schemes are described in further detail below. Other scheduling schemes that can provide good or near-optimum performance, and which may require less processing and/or statistic, may also be used, and this is within the scope of the invention.

[1070] FIG. 2 is a flow diagram of a process 200 to schedule terminals for downlink data transmission. Process 200 may be used to implement various downlink scheduling schemes, as described below. For clarity, the overall process is described first, and the details for some of the steps in the process are described subsequently.

[1071] In an embodiment, the transmission channels are assigned to the active terminals by evaluating one frequency subchannel group at a time. The first frequency subchannel group is considered by setting the frequency index $k = 1$, at step 210. The spatial subchannels for the k-th frequency subchannel group are then assigned to the terminals for downlink transmission starting at step 212. For the N-SIMO and MIMO modes on the downlink, assignment of spatial subchannels to the terminals is equivalent to assignment of the base station's transmit antennas to the terminals, since it is assumed that $N_S = N_T$.

[1072] Initially, one or more performance metrics to be used to select the best set of terminals for downlink transmission are initialized, at step 212. Various performance metrics may be used to evaluate the terminal sets and some of these are described in

further detail below. For example, a performance metric that maximizes system throughput may be used.

[1073] A new set of one or more active terminals is then selected from among all active terminals and considered for transmit antenna assignment, at step 214. This set of terminals forms a hypothesis to be evaluated. Various techniques may be used to limit the number of active terminals to be considered for scheduling, which then reduces the number of hypotheses to be evaluated, as described below. For each terminal in the hypothesis, the SNR vector, $\underline{\gamma}_i(k) = [\gamma_{i,1}(k), \gamma_{i,2}(k), ..., \gamma_{i,N_T}(k)]$, indicative of the post-processed SNRs for the N_T transmit antennas in the k-th frequency subchannel group is retrieved, at step 216. For the MIMO mode, a single MIMO terminal is selected for evaluation for the k-th frequency subchannel group, and one SNR vector for this terminal is retrieved. For the N-SIMO mode, N_T SIMO terminals are selected for evaluation, and N_T SNR vectors for these terminals are retrieved. And for the mixed mode, SNR vectors are retrieved for the SIMO and MIMO terminals in the selected set. For each MIMO terminal in the MIMO and mixed modes, the SNR vector may be replicated (or appropriately marked) such that the number of SNR vectors for this terminal is equal to the number of transmit antennas to be assigned to the terminal. The SNR vectors for all selected terminals in the hypothesis are used to form the hypothesis matrix $\underline{\Gamma}(k)$ shown in equation (2).

[1074] For each hypothesis matrix $\underline{\Gamma}(k)$ for N_T transmit antennas and N_T terminals, there are N_T factorial possible combinations of assignments of transmit antennas to terminals (i.e., $N_T!$ sub-hypotheses). Since a MIMO terminal is represented as multiple terminals in the matrix $\underline{\Gamma}(k)$, fewer sub-hypotheses exist if the hypothesis matrix $\underline{\Gamma}(k)$ includes one or more MIMO terminals. In any case, a particular new combination of antenna/terminal assignments is selected for evaluation, at step 218. This combination includes one antenna assigned to each of the N_T terminals. The antenna assignment may be performed such that all possible combinations of antenna/terminal assignments are eventually evaluated. Alternatively, a specific scheme may be used to assign antennas to the terminals, as described below. The new combination of antenna/terminal assignments forms a sub-hypothesis to be evaluated.

[1075] The sub-hypothesis is then evaluated and the performance metric (e.g., the system throughput) corresponding to this sub-hypothesis is determined (e.g., based on the SNRs for the sub-hypothesis), at step 220. The performance metric corresponding

to the best sub-hypothesis is then updated to reflect the performance metric for the current sub-hypothesis, at step 222. Specifically, if the performance metric for the current sub-hypothesis is better than that for the best sub-hypothesis, then the current sub-hypothesis becomes the new best sub-hypothesis, and the performance metric, terminal metrics, and antenna/terminal assignments corresponding to this sub-hypothesis are saved. The performance and terminal metrics are described below.

[1076] A determination is then made whether or not all sub-hypotheses for the current hypothesis have been evaluated, at step 224. If all sub-hypotheses have not been evaluated, then the process returns to step 218 and a different and not yet evaluated combination of antenna/terminal assignments is selected for evaluation. Steps 218 through 224 are repeated for each sub-hypothesis to be evaluated.

[1077] If all sub-hypotheses for the current hypothesis have been evaluated, at step 224, then a determination is next made whether or not all hypotheses have been considered for the current frequency subchannel group, at step 226. If all hypotheses have not been considered, then the process returns to step 214 and a different and not yet considered set of terminals is selected for evaluation. Steps 214 through 226 are repeated for each hypothesis to be considered for the current frequency subchannel group.

[1078] If all hypotheses for the current frequency subchannel group have been evaluated, at step 226, then the results for the best sub-hypothesis for this frequency subchannel group are saved, at step 228. The best sub-hypothesis corresponds to a specific set of one or more active terminals that provides the best performance for the frequency subchannel group.

[1079] If the scheduling scheme requires other system and terminal metrics to be maintained (e.g. the average throughput over the past N_P time slots, latency for data transmission, and so on), then these metrics are updated and may be saved, at step 230. The terminal metrics may be used to evaluate the performance of the individual terminals, and are described below.

[1080] A determination is then made whether or not all frequency subchannels have been assigned for downlink transmission, at step 232. If all frequency subchannels have not been assigned, then the next frequency subchannel group is considered by incrementing the index k (i.e., $k = k + 1$), at step 234. The process then returns to step 212 to assign the spatial subchannels of this new frequency subchannel group to the

terminals for downlink transmission. Steps 212 through 234 are repeated for each frequency subchannel group to be assigned.

[1081] If all frequency subchannel groups have been assigned, at step 232, then the data rates and the coding and modulation schemes for the terminals in the best sub-hypothesis for each frequency subchannel group are determined (e.g., based on their post-processed SNRs), at step 236. A schedule indicative of the specific active terminals selected for downlink data transmission, their assigned transmission channels, the scheduled time slot(s), the data rates, the coding and modulation schemes, other information, or any combination thereof, may be formed and communicated to these terminals (e.g., via a control channel) prior to the scheduled time slot, also at step 236. Alternatively, the active terminals may perform "blind" detection and attempt to detect all transmitted data streams to determine which ones, if any, of the data streams are intended for them. The downlink scheduling is typically performed for each scheduling interval, which may correspond to one or more time slots.

[1082] The process shown in FIG. 2 may be used to implement the various downlink scheduling schemes described above. For the exhaustive scheduling scheme, each available transmission channel may be assigned to any active terminal. This may be achieved by considering (1) all possible sets of terminals (i.e., all possible hypotheses) for each frequency subchannel group and (2) all possible antenna assignments for each terminal set (i.e., all possible sub-hypotheses). This scheme may provide the best performance and most flexibility, but also requires the most processing to schedule terminals for downlink data transmission.

[1083] For the priority-based scheduling scheme, the active terminals to be considered for assignment of transmission channels may be selected based on their priority, and the performance metric may also be made a function of the terminal priority, as described below. This scheme can reduce the number of terminals to be considered for transmission channel assignment, which then reduces scheduling complexity. For the FDM-TDM scheduling scheme, one MIMO terminal is assigned all of the spatial subchannels for each frequency subchannel group. In this case, the hypothesis matrix $\underline{\Gamma}(k)$ includes a single vector of post-processed SNRs for one MIMO terminal, and there is only one sub-hypothesis for each hypothesis. And for the SDMA-TDM scheduling scheme, all frequency subchannels of each spatial subchannel are

assigned to a single terminal, which may be a SIMO or MIMO terminal. For this scheme, steps 210, 212, 232, and 234 in FIG. 2 may be omitted.

[1084] For a given hypothesis matrix $\underline{\Gamma}(k)$, the scheduler evaluates various combinations of transmit antenna and terminal pairings (i.e., sub-hypotheses) to determine the best antenna/terminal assignments for the hypothesis. Various schemes may be used to assign transmit antennas to the terminals to achieve various system goals such as fairness, high performance, and so on.

[1085] In one antenna assignment scheme, all possible sub-hypotheses are evaluated based on a particular performance metric, and the sub-hypothesis with the best performance metric is selected. For each hypothesis matrix $\underline{\Gamma}(k)$, there are N_T factorial (i.e., $N_T!$) possible sub-hypotheses that may be evaluated. Each sub-hypothesis corresponds to a specific assignment of each transmit antenna to a particular terminal. Each sub-hypothesis may thus be represented with a vector of post-processed SNRs, which may be expressed as:

$$\underline{\gamma}_{\text{sub-hyp}}(k) = \{\gamma_{a,1}(k),\ \gamma_{b,2}(k),\ \ldots,\ \gamma_{r,N_T}(k)\} \ , \qquad \text{Eq (3)}$$

where $\gamma_{i,j}(k)$ is the post-processed SNR for the data stream from the j-th transmit antenna to the i-th terminal for the k-th frequency subchannel group, and the subscripts $\{a, b, \ldots$ and $r\}$ identify the specific terminals in the transmit antenna/terminal pairings for the sub-hypothesis.

[1086] Each sub-hypothesis is further associated with a performance metric, $R_{\text{sub-hyp}}(k)$, which may be a function of various factors. For example, a performance metric based on the post-processed SNRs may be expressed as:

$$R_{\text{sub-hyp}}(k) = f(\underline{\gamma}_{\text{sub-hyp}}(k)) \ , \qquad \text{Eq (4)}$$

where $f(\cdot)$ is a particular positive real function of the argument(s) within the parenthesis.

[1087] Various functions may be used to formulate the performance metric. In one embodiment, a function of the achievable throughput for all N_T transmit antennas for the sub-hypothesis may be used as the performance metric, which may be expressed as:

$$f(\underline{\gamma}_{sub-hyp}(k)) = \sum_{j=1}^{N_T} r_j(k) \ , \qquad \text{Eq (5)}$$

where $r_j(k)$ is the throughput associated with the j-th transmit antenna in the sub-hypothesis for the k-th frequency subchannel group, and may be expressed as:

$$r_j(k) = c_j \cdot \log_2(1 + \gamma_j(k)) \ , \qquad \text{Eq (6)}$$

where c_j is a positive constant that reflects the fraction of the theoretical capacity achieved by the coding and modulation scheme selected for the data stream transmitted on the j-th transmit antenna, and $\gamma_j(k)$ is the post-processed SNR for the j-th data stream on the k-th frequency subchannel group.

[1088] To simplify the scheduling, the resource allocation may be performed based on groups of multiple frequency subchannels instead of groups of single frequency subchannels. Even if a given group includes multiple frequency subchannels, the frequency selective nature of the channel response may be considered in allocating resources to the terminals. This may be achieved by evaluating the performance metric based on the response for the group of frequency subchannels. For example, the resource allocation may be performed based on groups of N_k frequency subchannels, where $N_k \geq 2$. The channel response over the N_k frequency subchannels may then be used to evaluate the performance metric. If the performance metric is throughput, then the summation of the achievable rates in equation (5) may be performed over both transmit antennas and frequency subchannels, as follows:

$$f(\underline{\gamma}_{sub-hyp}(k)) = \sum_{j=1}^{N_T} \sum_{i=1}^{N_k} r_j(i) \ ,$$

where $r_j(i)$ is the throughput associated with the j-th transmit antenna in the sub-hypothesis for the i-th frequency subchannel, and N_k is the number of frequency subchannels for the k-th frequency subchannel group. Thus, even if scheduling and resource allocation are performed for groups of multiple frequency subchannels, the performance of the individual frequency subchannels in each group may be considered in the scheduling.

[1089] The first antenna assignment scheme described above and used in FIG. 2 represents a specific scheme that can evaluate all possible combinations of assignments of transmit antennas to terminals. The number of potential sub-hypotheses to be evaluated by the scheduler for each hypothesis may be as large as $N_T!$, which may then result in a large number of total sub-hypotheses to be evaluated since a large number of hypotheses may also be considered.

[1090] The scheduling scheme shown in FIG. 2 performs an exhaustive search to determine the sub-hypothesis that provides the "optimal" system performance, as quantified by the performance metric used to select the best sub-hypothesis. A number of techniques may be used to reduce the complexity of the processing to assign transmit antennas to the terminals. One of these techniques is described below, and others may also be used and are within the scope of the invention. These techniques may also provide high system performance while reducing the amount of processing required to assign antennas to terminals.

[1091] In a second antenna assignment scheme, a maximum-maximum ("max-max") criterion is used to assign transmit antennas to the terminals in the hypothesis being evaluated. Using this max-max criterion, each transmit antenna is assigned to the terminal that achieves the best SNR for the transmit antenna. The antenna assignment may be performed for each frequency subchannel group and for one transmit antenna at a time.

[1092] FIG. 3 is a flow diagram of a process 218a to assign transmit antennas to terminals for a particular frequency subchannel group using the max-max criterion. Process 218a is performed for a particular hypothesis, which corresponds to a specific set of one or more active terminals to be evaluated. Process 218a may be used for step 218 in FIG. 2, in which case only one sub-hypothesis is evaluated for each hypothesis in process 200.

[1093] Initially, the maximum SNR in the hypothesis matrix $\underline{\Gamma}(k)$ is determined, at step 312. This maximum SNR corresponds to a specific transmit antenna/terminal pairing, and the transmit antenna is assigned to this terminal, at step 314. This transmit antenna and terminal are then removed from the matrix $\underline{\Gamma}(k)$, and the matrix is reduced to dimension $(N_T-1)\times(N_T-1)$ by removing both the column corresponding to the transmit antenna and the row corresponding to the terminal just assigned, at step 316.

[1094] At step 318, a determination is made whether or not all transmit antennas in the hypothesis have been assigned. If all transmit antennas have been assigned, then the antenna assignments are provided, at step 320, and the process terminates. Otherwise, the process returns to step 312 and another transmit antenna is assigned in similar manner.

[1095] Table 1 shows an example matrix $\underline{\Gamma}(k)$ of SNRs derived by the terminals in a 4x4 MIMO system in which the base station includes four transmit antennas and each terminal includes four receive antennas. For the antenna assignment scheme based on the max-max criterion, the best SNR (16 dB) in the original 4x4 matrix is achieved by transmit antenna 3 and is assigned to terminal 1, as indicated by the shaded box in the third row of the fourth column in the table. Transmit antenna 3 and terminal 1 are then removed from the matrix. The best SNR (14 dB) in the reduced 3x3 matrix is achieved by both transmit antennas 1 and 4, which are respectively assigned to terminals 3 and 2. The remaining transmit antenna 2 is then assigned to terminal 4.

Table 1

SNR (dB)	Transmit Antenna			
Terminal	1	2	3	4
1	7	9	16	5
2	8	10	12	14
3	14	7	6	9
4	12	10	7	5

[1096] Table 2 shows the antenna assignments using the max-max criterion for the example matrix $\underline{\Gamma}(k)$ shown in Table 1. For terminal 1, the best SNR (16 dB) is achieved when processing the signal transmitted from transmit antenna 3. The best transmit antennas for the other terminals are also indicated in Table 2. The scheduler may then use this information to select the proper coding and modulation scheme to use for data transmission.

Table 2

Terminal	Transmit Antenna	SNR (dB)
1	3	16
2	4	14
3	1	14
4	2	10

[1097] Once the antenna assignments have been made for a particular hypothesis matrix $\underline{\Gamma}(k)$ using the max-max criterion, the performance metric (e.g., the system throughput) corresponding to this hypothesis may be determined (e.g., based on the SNRs corresponding to the antenna assignments), as shown in equations (4) through (6). This performance metric is updated for each hypothesis in a particular frequency subchannel group. When all hypotheses for the frequency subchannel group have been evaluated, the best set of terminals and antenna assignments are selected for downlink data transmission on the frequency subchannel group in an upcoming time slot. The scheduling may be performed for each of the N_G frequency subchannel groups.

[1098] The downlink scheduling scheme described in FIGS. 2 and 3 represents a specific scheme that evaluates various hypotheses corresponding to various possible sets of active terminals (which may include SIMO and/or MIMO terminals) desiring downlink data transmission in an upcoming time slot. The total number of hypotheses to be evaluated by the scheduler can be quite large, even for a small number of active terminals. In fact, the total number of hypotheses, N_{hyp}, can be expressed as:

$$N_{hyp} = N_G \cdot \binom{N_U}{N_T} = \frac{N_G \cdot N_U!}{(N_U - N_T)!N_T!} , \qquad \text{Eq (7)}$$

where N_U is the number of active terminals to be considered for scheduling. For example, if $N_G = 16$, $N_U = 8$, and $N_T = 4$, then $N_{hyp} = 1120$. An exhaustive search may be used to determine the particular hypothesis and the particular antenna assignments that provide the best system performance, as quantified by the performance metric used to select the best hypothesis and antenna assignments.

[1099] As noted above, other downlink scheduling schemes having reduced complexity may also be implemented. These scheduling schemes may also provide

high system performance while reducing the amount of processing required to schedule terminals for downlink data transmission.

[1100] In the priority-based scheduling scheme, active terminals are scheduled for data transmission based on their priority. The priority of each active terminal may be derived based on one or more metrics (e.g., average throughput), system constraints and requirements (e.g., maximum latency), other factors, or a combination thereof, as described below. A list may be maintained for all active terminals desiring data transmission in an upcoming time slot. When a terminal desires downlink data transmission, it is added to the list and its metrics are initialized (e.g., to zero). The metrics of each terminal in the list are thereafter updated at each time slot. Once a terminal no longer desires data transmission, it is removed from the list.

[1101] For each frequency subchannel group in each time slot, all or a subset of the terminals in the list may be considered for scheduling. The specific number of terminals to be considered may be selected based on various factors. In one embodiment, only the N_T highest priority terminals are selected for data transmission. In another embodiment, the highest N_X priority terminals in the list are considered for scheduling, where $N_X > N_T$. A MIMO terminal may be represented as multiple terminals when selecting the N_T or N_X highest priority terminals for scheduling. For example, if $N_T = 4$ and four independent data streams are transmitted from the base station for a given frequency subchannel group, then one SIMO terminal may be selected along with a MIMO terminal to be assigned three spatial subchannels (in which case the MIMO terminal is effectively representing three terminals in selecting the four highest priority terminals).

[1102] FIG. 4 is a flow diagram for a priority-based downlink scheduling scheme 400 whereby a set of N_T highest priority terminals is considered for scheduling for each frequency subchannel group. Initially, the first frequency subchannel group is considered by setting the frequency index $k = 1$, at step 410. The spatial subchannels for the k-th frequency subchannel group are then assigned to the terminals for downlink transmission starting at step 412.

[1103] The scheduler examines the priority for all active terminals in the list and selects the set of N_T highest priority terminals, at step 412. The remaining active terminals in the list are not considered for scheduling for this frequency subchannel group in this scheduling interval. The channel estimates for each selected terminal are

then retrieved, at step 414. For example, the post-processed SNRs for the N_T selected terminals may be retrieved and used to form the hypothesis matrix $\underline{\Gamma}(k)$.

[1104] The N_T transmit antennas are then assigned to the N_T selected terminals based on the channel estimates and using any one of a number of antenna assignment schemes, at step 416. For example, the antenna assignments may be based on an exhaustive search or the max-max criterion described above. In another antenna assignment scheme, the transmit antennas are assigned to the terminals such that their priorities are normalized as close as possible, after the terminal metrics are updated.

[1105] The data rates and the coding and modulation schemes for the terminals are then determined based on the antenna assignments, at step 418. The metrics of the scheduled (and unscheduled) terminals in the list are updated to reflect the scheduled data transmission (and non-transmission, respectively), and the system metrics are also updated, at step 420.

[1106] A determination is then made whether or not all frequency subchannels have been assigned for downlink transmission, at step 422. If all frequency subchannels have not been assigned, then the next frequency subchannel group is considered by incrementing the index k (i.e., $k = k + 1$), at step 424. The process then returns to step 412 to assign the spatial subchannels of this new frequency subchannel group to the same or a different set of active terminals. Steps 412 through 424 are repeated for each frequency subchannel group to be assigned.

[1107] If all frequency subchannel groups have been assigned, at step 422, then a schedule indicative of the specific active terminals selected for downlink data transmission, their assigned transmission channels, the scheduled time slot(s), the data rates, the coding and modulation schemes, and so on, or any combination thereof, may be formed and communicated to these terminals, at step 426. The process then terminates for this scheduling interval.

[1108] As noted above, the transmit antennas may be assigned to the selected terminals for each frequency subchannel group based on various schemes. In one antenna assignment scheme, the transmitted antennas are assigned to achieve high system performance and based on the priority of the terminals.

[1109] Table 3 shows an example of the post-processed SNRs derived by each terminal in a hypothesis being considered, which is for a specific frequency subchannel group. For terminal 1, the best SNR is achieved when detecting the data stream

transmitted from transmit antenna 3, as indicated by the shaded box in row 3, column 4 of the table. The best transmit antennas for other terminals in the hypothesis are also indicated by the shading in the boxes.

Table 3

SNR (dB)	Transmit Antenna			
Terminal	1	2	3	4
1	7	9	16	5
2	8	10	12	14
3	14	7	6	9
4	12	10	7	5

[1110] If each terminal identifies a different transmit antenna from which the best post-processed SNR is detected, then the transmit antennas may be assigned to the terminals based on their best post-processed SNRs. For the example shown in Table 3, terminal 1 may be assigned to transmit antenna 3, and terminal 2 may be assigned to transmit antenna 2.

[1111] If more than one terminal prefers the same transmit antenna, then the scheduler can determine the antenna assignments based on various criteria (e.g., fairness, performance metric, and others). For example, Table 3 indicates that the best post-processed SNRs for terminals 3 and 4 occur for the data stream transmitted from the same transmit antenna 1. If the objective is to maximize throughput, then the scheduler may assign transmit antenna 1 to terminal 3 and transmit antenna 2 to terminal 4. However, if antennas are assigned to achieve fairness, then transmit antenna 1 may be assigned to terminal 4 if terminal 4 has higher priority than terminal 3.

[1112] The scheduling for MIMO terminals may also be performed based on full-CSI. In this case, the statistic to be used for scheduling terminals is the complex channel gains between the base station's transmit antennas and the terminal's receive antennas, which are used to form the channel response matrix, $\underline{\mathbf{H}}(k)$, shown in equation (1). The scheduling is then performed such that a set of mutually compatible spatial signatures is selected for each frequency subchannel group. Scheduling of terminals based on the channel response matrix, $\underline{\mathbf{H}}(k)$, is described in further detail below.

Downlink Scheduling for MISO Terminals

[1113] For the N-MISO mode, where ($N_R < N_T$), complex channel gains between the transmit antennas at the base station and the receive antenna(s) at the terminals may be used to form the channel response matrix, $\underline{\mathbf{H}}(k)$, shown in equation (1) for each set of MISO terminals to be evaluated. The selection of MISO terminals for downlink transmission is then performed over the active terminals, and the selection goal is mutually compatible spatial signatures over the band of interest.

[1114] For the downlink in the multi-user N-MISO mode, the base station employs N_T transmit antennas and (for simplicity) each of the N_U MISO terminals to be considered for downlink scheduling employs a single receive antenna (i.e., $N_R = 1$). In this case, up to N_T terminals may be served by the base station simultaneously on any given frequency subchannel group (i.e., $N_U \le N_T$). The model of the MISO channel for terminal i may be expressed as:

$$y_i(k) = \underline{\mathbf{H}}_i(k)\underline{\mathbf{x}}(k) + n_i(k) \quad , \qquad \text{Eq (8)}$$

where $y_i(k)$ is the symbol received by the i-th terminal, for $i \in \{1, ..., N_U\}$, on the k-th frequency subchannel group;

$\underline{\mathbf{x}}(k)$ is the transmitted vector (i.e., $\underline{\mathbf{x}} = [x_1 \; x_2 \ldots x_{N_T}]^T$), where $\{x_j\}$ is the entry transmitted from the j-th transmit antenna for $j \in \{1, ..., N_T\}$, and for any matrix, $\underline{\mathbf{M}}$, $\underline{\mathbf{M}}^T$ denotes the transpose of $\underline{\mathbf{M}}$;

$\underline{\mathbf{H}}_i(k)$ is the $1 \times N_T$ channel response vector for the MISO channel of the i-th terminal for the k-th frequency subchannel group, where element $h_{i,j}$ is the coupling (i.e., the complex gain) between the j-th transmit antenna and the receive antenna of the i-th terminal, for $i \in \{1, ..., N_U\}$ and $j \in \{1, ..., N_T\}$; and

$n_i(k)$ is the additive white Gaussian noise (AWGN) for the k-th frequency subchannel group of the i-th terminal, which has a mean of 0 and a variance of σ_i^2.

[1115] For simplicity, each frequency subchannel group is assumed to be a flat-fading, narrowband channel that can be represented by a constant complex value.

Therefore, the elements of the channel response vector, $\underline{\mathbf{H}}_i(k)$, for $i \in \{1, ..., N_U\}$, are scalars. In addition, it is assumed that there is a maximum power limit on each transmit antenna, which is denoted as $P_{max,j}$, for $j \in \{1, ..., N_T\}$. The transmit power on antenna j at any given time is denoted as P_j, where $P_j \le P_{max,j}$.

[1116] The N_T data streams transmitted from the N_T transmit antennas for each frequency subchannel group can interfere with each other at the receive antenna of each terminal according to the channel response vectors, $\underline{\mathbf{H}}_i(k)$. Without any pre-processing at the base station, the different data streams intended for different MISO terminals are subject to interference, which is referred to as multi-access interference (MAI). Because each MISO terminal employs only one receive antenna, all spatial processing aimed at combating the channel and MAI needs to be performed at the transmitter.

[1117] If the base station has knowledge of the channel response vector, $\underline{\mathbf{H}}_i(k)$, for each MISO terminal to be considered for downlink scheduling (i.e., full channel state information), one technique for eliminating or reducing the MAI is by use of channel correlation matrix inversion (CCMI).

[1118] The transmit vector at the base station is $\underline{\mathbf{x}}(k) = [x_1(k) \; x_2(k) \; ... \; x_{N_T}(k)]^T$, where $\{x_j(k)\}$ is the entry transmitted from the j-th transmit antenna for the k-th frequency subchannel group. Denoting the data stream intended for terminal i by $d_i(k)$, the actual data vector is $\underline{\mathbf{d}}(k) = [d_1(k) \; d_2(k) \; ... \; d_{N_U}(k)]^T$, where the relationship between the data vector and the transmitted vector is may be expressed as:

$$\underline{\mathbf{x}}(k) = \underline{\mathbf{A}}(k)\underline{\mathbf{S}}(k)\underline{\mathbf{d}}(k) \quad , \qquad \text{Eq (9)}$$

where $\underline{\mathbf{A}}(k)$ is an $N_T \times N_U$ CCMI matrix and $\underline{\mathbf{S}}(k)$ is an $N_U \times N_U$ scaling matrix. The CCMI matrix may be viewed as including a number of steering vectors, one for each MISO terminal, with each steering vector being used to generate a beam for a respective MISO terminal. The CCMI technique decorrelates the data streams for the MISO terminals, and the solution for $\underline{\mathbf{A}}(k)$ may be expressed as:

$$\underline{\mathbf{A}}(k) = \underline{\mathbf{H}}^T(k)\left(\underline{\mathbf{H}}(k)\underline{\mathbf{H}}^T(k)\right)^{-1} \quad , \qquad \text{Eq (10)}$$

where $\underline{\mathbf{H}}(k) = \begin{bmatrix} \underline{\mathbf{H}}_1(k) \\ \underline{\mathbf{H}}_2(k) \\ \mathrm{M} \\ \underline{\mathbf{H}}_{N_U}(k) \end{bmatrix}$ is an $\mathrm{N_U} \times \mathrm{N_T}$ matrix that holds the channel response vectors of the set of $\mathrm{N_U}$ MISO terminals being considered for downlink scheduling for the current hypothesis.

[1119] The solution for $\underline{\mathbf{A}}(k)$ does not require $\underline{\mathbf{H}}(k)$ to be a square matrix, which is the case when $\mathrm{N_U} \neq \mathrm{N_T}$. However, if $\underline{\mathbf{H}}(k)$ is a square matrix, then the solution in equation (10) can be rewritten as $\underline{\mathbf{A}}(k) = \underline{\mathbf{H}}^{-1}(k)$, where $\underline{\mathbf{H}}^{-1}(k)$ is the inverse of $\underline{\mathbf{H}}(k)$, so that $\underline{\mathbf{H}}^{-1}(k)\underline{\mathbf{H}}(k) = \underline{\mathbf{H}}(k)\underline{\mathbf{H}}^{-1}(k) = \underline{\mathbf{I}}$, where $\underline{\mathbf{I}}$ is the square identity matrix with ones on the diagonal and zeros elsewhere.

[1120] Because there is a power limit of $P_{max,j}$ on each transmit antenna $j \in \{1, \ldots, \mathrm{N_T}\}$, it may be necessary to scale the rows of $\underline{\mathbf{A}}(k)$ to ensure that the power used on transmit antenna j, P_j, does not exceed $P_{max,j}$. However, in order to maintain the orthogonality between the rows of $\underline{\mathbf{H}}(k)$ and the columns of $\underline{\mathbf{A}}(k)$, all entries within each column of $\underline{\mathbf{A}}(k)$ need be scaled by the same value. The scaling is accomplished by the scaling matrix, $\underline{\mathbf{S}}(k)$, in equation (9), which has the following form:

$$\underline{\mathbf{S}}(k) = \begin{bmatrix} S_1(k) & 0 & \mathrm{A} & 0 \\ 0 & S_2(k) & \mathrm{A} & 0 \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ 0 & 0 & \mathrm{A} & S_{N_U}(k) \end{bmatrix}, \qquad \text{Eq (11)}$$

where the scale value $S_i(k)$ multiplies data stream $d_i(k)$. The set of scale values, $\{S_i(k)\}$, can be obtained by solving the following set of equations

$$\mathrm{diag}\left((\underline{\mathbf{A}}(k)\underline{\mathbf{S}}(k))(\underline{\mathbf{A}}(k)\underline{\mathbf{S}}(k))^T \right) \leq \underline{\mathbf{P}}_{max}(k) , \qquad \text{Eq (12)}$$

where $\underline{\mathbf{P}}_{max}(k) = [P_{max,1}(k) \quad P_{max,2}(k) \quad \mathrm{A} \quad P_{max,N_T}(k)]^T$ and $P_{max,j}(k)$ is the maximum power allocated to the k-th frequency subchannel group for the j-th transmit antenna. The values $S_i(k)$ can be solved for from equation (12) and ensure that the power used

on each transmit antenna for the k-th frequency subchannel group does not exceed $P_{\max,j}(k)$.

[1121] The total transmit power, $P_{\max,j}$, for each transmit antenna may be allocated to the N_G frequency subchannel groups in various manners. In one embodiment, the total transmit power is allocated equally among the N_G frequency subchannel groups such that $P_j(k) = P_{\max,j} / N_G$. In another embodiment, the total transmit power can be allocated unequally among the N_G frequency subchannel groups while maintaining $\sum_{k=1}^{N_G} P_j(k) = P_{\max,j}$. The total transmit power, $P_{\max,j}$, may be allocated based on various techniques, including a "water-pouring" or "water-filling" technique that allocates transmit power such that throughput is maximized. The water-pouring technique is described by Robert G. Gallager in "Information Theory and Reliable Communication," John Wiley and Sons, 1968, which is incorporated herein by reference. A specific algorithm for performing the basic water-pouring process for a MIMO-OFDM system is described in U.S. Patent Application Serial No. 09/978,337, entitled "Method and Apparatus for Determining Power Allocation in a MIMO Communication System," filed October 15, 2001, assigned to the assignee of the present application and incorporated herein by reference. Transmit power allocation is also described in U.S. Patent Application Serial No. 10/017,308 entitled "Time-Domain Transmit and Receive Processing with Channel Eigen-mode Decomposition for MIMO Systems," filed December 7, 2001, assigned to the assignee of the present application and incorporated herein by reference. The optimum allocation of the total transmit power, $P_{\max,j}$, for each of the N_T transmit antennas among the N_G frequency subchannel groups is typically complex, and iterative techniques may be used to solve for the optimum power allocation.

[1122] Substituting equation (9) into equation (8), the expression for the received symbol for terminal i may be expressed as:

$$y_i(k) = \underline{\mathbf{H}}_i(k)\underline{\mathbf{A}}(k)\underline{\mathbf{S}}(k)\underline{\mathbf{d}}(k) + n_i(k) \quad , \qquad \text{Eq (13)}$$

which simplifies to

$$y_i(k) = S_i(k)d_i(k) + n_i(k) \quad , \qquad \text{Eq (14)}$$

since $\underline{\mathbf{H}}_i(k)$ is orthogonal to all, except the i-th, columns of $\underline{\mathbf{A}}(k)$.

[1123] The resulting SNR for terminal i for the k-th frequency subchannel group may be expressed as:

$$\gamma_i(k) = \frac{S_i^2(k)}{\sigma_i^2(k)} \ . \qquad \text{Eq (15)}$$

[1124] In selecting a set of MISO terminals having mutually compatible spatial signatures for downlink data transmission on a given frequency subchannel group, the above analysis may be performed for each set of MISO terminals to be evaluated (i.e., each hypothesis). The SNR for each terminal in the set may be determined as shown in equation (15). This SNR may be used in a performance metric, such as the one based on throughput shown above in equations (5) and (6). Mutual compatibility may thus be defined based on throughput or some other criteria (e.g., the most mutually compatible MISO terminals may be the ones that achieve the highest overall throughput).

[1125] The MISO terminals may also be scheduled for downlink data transmission based on their priorities. In this case, the above description for scheduling SIMO and MIMO terminals based on priority may also be applied for scheduling MISO terminals. For example, the N_T highest priority MISO terminals may be considered for scheduling for each frequency subchannel group.

[1126] Other techniques to generate multiple beams for multiple terminals may also be used, and this is within the scope of the invention. For example, the beam steering may be performed based on a minimum mean square error (MMSE) technique. The CCMI and MMSE techniques are described in detail in U.S. Patent Application Serial No. U.S. Patent Application Serial Nos. 09/826,481 and 09/956,449, both entitled "Method and Apparatus for Utilizing Channel State Information in a Wireless Communication System," respectively filed March 23, 2001 and September 18, 2001, both assigned to the assignee of the present application and incorporated herein by reference.

[1127] Data transmission to multiple terminals concurrently based on spatial signatures is also described in U.S. Patent No. 5,515,378, entitled "Spatial Division Multiple Access Wireless Communication System," issued May 7, 1996, which is incorporated herein by reference.

[1128] The beam-steering technique described above for MISO terminals may also be used for MIMO terminals.

[1129] The ability to schedule MISO terminals on a per frequency subchannel group basis can result in improved system performance since the frequency signatures of the MISO terminals may be exploited in selecting the set of mutually compatible terminals for each frequency subchannel group.

[1130] The techniques described above may be generalized to handle a combination of SIMO, MISO, and MIMO terminals. For example, if four transmit antennas are available at the base station, then four independent data streams may be transmitted to a single 4x4 MIMO terminal, two 2x4 MIMO terminals, four 1x4 SIMO terminals, four 4x1 MISO terminals, one 2x4 MIMO terminal and two 1x4 SIMO terminals, or any other combination of terminals designated to receive a total of four data streams for each frequency subchannel group. The scheduler can be designed to select the best combination of terminals based on the post-processed SNRs for various hypothesized sets of terminals, where each hypothesized set may include a mixture of SIMO, MISO, and MIMO terminals.

[1131] Various metrics and factors may be used to determine the priority of the active terminals. In an embodiment, a "score" may be maintained for each active terminal and for each metric to be used for scheduling. In one embodiment, a score indicative of an average throughput over a particular averaging time interval is maintained for each active terminal. In one implementation, the score $\phi_i(n)$ for terminal i at time slot n is computed as a linear average throughput achieved over N_P prior time slots, and may be expressed as:

$$\phi_i(n) = \frac{1}{N_P} \sum_{\lambda=n-N_P+1}^{n} r_i(\lambda) / r_{max} \ , \qquad \text{Eq (16)}$$

where $r_i(n)$ is the "realized" data rate (in unit of bits/time slot) for terminal i at time slot n and may be computed based on the post-processed SNRs as shown in equation (6). Typically, $r_i(n)$ is bound by a particular maximum achievable data rate, r_{max}, and a particular minimum data rate (e.g., zero). In another implementation, the score $\phi_i(n)$ for terminal i at time slot n is an exponential average throughput achieved over some time interval, and may be expressed as:

$$\phi_i(n) = (1-\alpha)\cdot\phi_i(n-1) + \alpha\cdot r_i(n)/r_{max} \ , \qquad \text{Eq (17)}$$

where α is a time constant for the exponential averaging, with a larger value for α corresponding to a shorter averaging time interval.

[1132] When a terminal desires data transmission, it is added to the active terminals list and its score is initialized to zero. The score for each active terminal in the list may subsequently be updated at each time slot. Whenever an active terminal is not scheduled for transmission in a given time slot, its data rate for the time slot is set to zero (i.e., $r_i(n)=0$) and its score is updated accordingly. If the data packet transmitted in a scheduled time slot is received in error by a terminal, then the terminal's effective data rate for that time slot may be set to zero. The packet error may not be known immediately (e.g., due to round trip delay of an acknowledgment/negative acknowledgment (Ack/Nak) scheme used for the data transmission) but the score can be adjusted accordingly once this information is available.

[1133] The priority for the active terminals may also be determined based in part on system constraints and requirements. For example, if the maximum latency for a particular terminal exceeds a threshold value, then the terminal may be elevated to a high priority.

[1134] Other factors may also be considered in determining the priority of the active terminals. One such factor may be related to the type of data to be transmitted to the terminals. Delay sensitive data may be associated with higher priority, and delay insensitive data may be associated with lower priority. Retransmitted data due to decoding errors in a prior transmission may also be associated with higher priority since other processes may be waiting at the terminal for the retransmitted data. Another factor may be related to the type of data service being provided for the terminals. Other factors may also be considered in determining priority and are within the scope of the invention.

[1135] The priority of an active terminal may thus be a function of any combination of (1) the score maintained for the terminal for each metric to be considered, (2) other parameter values maintained for system constraints and requirements, and (3) other factors. In one embodiment, the system constraints and requirements represent "hard" values (i.e., high or low priority, depending on whether or not the constraints and requirements have been violated) and the scores represent "soft" values. For this

embodiment, terminals for which the system constraints and requirements have not been met may be immediately considered, along with other terminals based on their scores.

[1136] A priority-based scheduling scheme may be designed to achieve equal average throughput (e.g., equal quality of service or QoS) for all active terminals in the list. In this case, the active terminals are prioritized based on their achieved average throughput, which may be determined as shown in equation (16) or (17). In this priority-based scheduling scheme, the scheduler uses the scores to prioritize terminals for assignment to the available transmission channels. The scores of the terminals are updated based on their assignments or non-assignments to transmission channels and may further be adjusted for packet errors. The active terminals in the list may be prioritized such that the terminal with the lowest score is given the highest priority, and the terminal with the highest score is conversely given the lowest priority. Other methods for ranking terminals may also be used. The prioritization may also assign non-uniform weighting factors to the terminal scores.

[1137] For a downlink scheduling scheme in which terminals are selected and scheduled for data transmission based on their priority, it is possible for poor terminal groupings to occur occasionally. A "poor" terminal set is one that results in similar channel response matrices $\underline{\mathbf{H}}(k)$ which cause poor SNRs for all terminals on all transmitted data streams. This then results in low throughput for each terminal in the set and low overall system throughput. When this occurs, the priorities of the terminals may not change substantially over several time slots. The scheduler may then be stuck with this particular terminal set until the priorities of the terminals change sufficiently to cause a change in membership in the set.

[1138] To avoid the "clustering" effect described above, the scheduler can be designed to recognize this condition prior to assigning terminals to the available transmission channels and/or to detect the condition once it has occurred. A number of schemes may be used to determine the degree of linear dependence in the channel response matrices $\underline{\mathbf{H}}(k)$. One scheme to detect clustering is to apply a particular threshold to the hypothesis matrix $\underline{\mathbf{\Gamma}}(k)$. If all or a substantial number of SNRs in the matrix $\underline{\mathbf{\Gamma}}(k)$ are below this threshold, then the clustering condition is deemed to be present. In the event that the clustering condition is detected, the scheduler can reorder the terminals (e.g., in a random manner) in an attempt to reduce the linear dependence in the hypothesis matrix. A shuffling scheme may also be devised to force the scheduler

to select terminal sets that result in "good" hypothesis matrices (i.e., ones that have minimal amount of linear dependence).

[1139] The scheduling of terminals for downlink data transmission and the scheduling of terminals based on priority are also described in U.S. Patent Application Serial No. 09/859,345, entitled "Method and Apparatus for Allocating Downlink Resources in a Multiple-Input Multiple-Output (MIMO) Communication System," filed May 16, 2001; U.S. Patent Application Serial No. 09/539,157, entitled "Method and Apparatus for Controlling Transmissions of a Communications System," filed March 30, 2000; and U.S. Patent Application Serial No. 09/675,706, entitled "Method and Apparatus for Determining Available Transmit Power in a Wireless Communication System," filed September 29, 2000, all assigned to the assignee of the present application and incorporated herein by reference.

[1140] Some of the downlink scheduling schemes described above employ techniques to reduce the amount of processing required to select terminals for evaluation and assign transmission channels to the selected terminals. These and other techniques may also be combined to derive other scheduling schemes, and this is within the scope of the invention. For example, the N_X highest priority terminals may be considered for scheduling using any one of the schemes described above.

[1141] For the downlink scheduling schemes described above, the total available transmit power for each transmit antenna is assumed to be allocated uniformly across all frequency subchannels selected for use for downlink data transmission. However, this uniform transmit power allocation is not a requirement. Other downlink scheduling schemes that select terminals for data transmission, assign transmission channels to the selected terminals, and further allocate transmit power to the assigned transmission channels may also be devised. Some of these scheduling schemes are described below.

[1142] In one downlink scheduling scheme with non-uniform transmit power allocation, only transmission channels with achieved SNRs above a particular threshold SNR are selected for use, and transmission channels with achieved SNRs below this threshold SNR are not used. This scheme may be used to remove poor transmission channels with limited transmission capabilities by allocating no transmit power to these transmission channels. The total available transmit power may then be allocated uniformly or non-uniformly across the selected transmission channels.

[1143] In another downlink scheduling scheme, the transmit power is allocated such that approximately equal SNRs are achieved for all transmission channels used to transmit each data stream. A particular data stream may be transmitted via multiple transmission channels (i.e., via multiple spatial subchannels and/or multiple frequency subchannels), and these transmission channels may achieve different SNRs if equal transmit power is allocated to these transmission channels. By allocating different amounts of transmit power to these transmission channels, approximately equal SNRs may be achieved which would then allow a single common coding and modulation scheme to be used for the data stream transmitted on these transmission channels. In effect, the unequal power allocation performs a channel inversion on the transmission channels such that they appear as being similar at the receiver. Channel inversion of all transmission channels and the channel inversion of only the selected transmission channels are described in U.S. Patent Application Serial No. 09/860,274, filed May 17, 2001, U.S. Patent Application Serial No. 09/881,610, filed June 14, 2001, and U.S. Patent Application Serial No. 09/892,379, filed June 26, 2001, all three entitled "Method and Apparatus for Processing Data for Transmission in a Multi-Channel Communication System Using Selective Channel Inversion," assigned to the assignee of the present application, and incorporated herein by reference.

[1144] In yet another downlink scheduling scheme, the transmit power may be allocated such that a desired data rate is achieved for each of the scheduled terminals. For example, more transmit power may be allocated to terminals with higher priority and less transmit power may be allocated to terminals with lower priority.

[1145] In yet another downlink scheduling scheme, the transmit power may be allocated non-uniformly to achieve high throughput. High system throughput may be achieved by allocating more transmit power to better transmission channels and less transmit power to poor transmission channels. The "optimum" allocation of transmit power to transmission channels of varying capacities may be performed based on the water-pouring technique. A scheme for allocating transmit power based on water pouring is described in the aforementioned U.S. Patent Application Serial No. 09/978,337.

[1146] Other downlink scheduling schemes that also allocate transmit power in a non-uniform manner to achieve the desired results may also be implemented, and this is within the scope of the invention.

[1147] Typically, the terminals determine their post-processed SNRs from some "assumed" power allocation, which may be the fixed power used for the pilot transmitted from the base station. Therefore, if the powers used for data transmission deviate from the assumed powers, then the post-processed SNRs will be different. Since the data rates used for the data transmission are based largely on the post-processed SNRs, the actual data rates may be sent to the terminals (e.g., in the preamble of a data packet). The terminals may also perform "blind" rate detection and attempt to process the received data transmission at various possible data rates until the data transmission is either received correctly or cannot be recovered error-free for all possible rates. Changing the transmit power in a given spatial subchannel can impact the post-processed SNR of another spatial subchannel in the same frequency subchannel group, and this effect can be considered in selecting terminals for data transmission.

[1148] "Water-filling" power allocation may also be used to allocate the available transmit power among the transmission channels such that throughput is maximized. The water-filling process may be performed in various manners such as (1) across all frequency subchannel groups for each spatial subchannel, (2) across all spatial subchannels for each frequency subchannel group, (3) across all frequency subchannels of all spatial subchannels, or (4) over some defined set of transmission channels. For example, the water-filling may be performed across a set of transmission channels used for a single data stream targeted for a particular terminal.

[1149] With partial-CSI schemes (e.g., those that use post-processed SNRs), there is a per antenna constraint on the allocation of transmit power. So for a multi-user case, the transmit powers may be allocated/reallocated (1) among multiple terminals scheduled on the same transmit antenna, (2) among the multiple transmission channels assigned to each scheduled terminal (with the total power allocated to each terminal being fixed), or (3) based on some other allocation scheme. For full-CSI schemes (e.g., those based on channel gains), additional flexibility is available since the transmit power may be reallocated across transmit antennas (i.e., eigenmodes) as well as across frequency subchannel groups. The allocation/reallocation of transmit power among multiple terminals then takes on an additional dimension.

[1150] Thus, more complex downlink scheduling schemes that may be able to achieve throughput closer to optimum may be devised. These scheduling schemes may evaluate a large number of hypotheses and antenna assignments (and possibly different

transmit power allocations) in order to determine the best set of terminals and the best antenna assignments. Other downlink scheduling schemes may also be designed to take advantage of the statistical distribution of the data rates achieved by each terminal. This information may be useful in reducing the number of hypotheses to evaluate. In addition, for some applications, it may be possible to learn which terminal groupings (i.e., hypotheses) work well by analyzing performance over time. This information may then be stored, updated, and used by the scheduler in future scheduling intervals.

[1151] The techniques described above may be used to schedule terminals for data transmission in the MIMO mode, the N-SIMO mode, and the mixed mode. Other considerations may also be applicable for each of these operating modes, as described below.

[1152] In the MIMO mode, (up to) N_T independent data streams may be simultaneously transmitted by the base station from N_T transmit antennas for each frequency subchannel group and targeted to a single MIMO terminal with N_R receive antennas (i.e., $N_R \times N_T$ MIMO). The MIMO terminal may use spatial equalization (for a non-dispersive MIMO channel with flat fading) or space-time equalization (for a dispersive MIMO channel with frequency selective fading) to process and separate the N_T transmitted data streams for each frequency subchannel group. The SNR of each post-processed data stream (i.e., after equalization) may be estimated and sent back to the base station as channel state information. The base station may then use this information to select the proper rate to use for each data stream such that the MIMO terminal is able to detect each transmitted data stream at the desired level of performance (e.g., the target PER).

[1153] If all data streams are transmitted to one terminal, as is the case in the MIMO mode, then the successive cancellation receiver processing technique may be used at this terminal to process N_R received signals to recover N_T transmitted data streams for each frequency subchannel group. This technique successively processes the N_R received signals a number of times (or iterations) to recover the signals transmitted from the base station, with one transmitted signal being recovered for each iteration. For each iteration, the technique performs spatial or space-time equalization on the N_R received signals. One of the transmitted signals is then recovered, and the interference due to the recovered signal is then estimated and canceled from the received signals to derive "modified" signals having the interference component removed.

[1154] The modified signals are then processed by the next iteration to recover another transmitted signal. By removing the interference due to each recovered signal from the received signals, the SNR improves for the transmitted signals included in the modified signals and not yet recovered. The improved SNR results in improved performance for the terminal as well as the system.

[1155] The successive cancellation receiver processing technique is described in further detail in U.S. Patent Application Serial No. 09/854,235, entitled "Method and Apparatus for Processing Data in a Multiple-Input Multiple-Output (MIMO) Communication System Utilizing Channel State Information," filed May 11, 2001, and U.S. Patent Application Serial No. 09/993,087, entitled "Multiple-Access Multiple-Input Multiple-Output (MIMO) Communication System," filed November 6, 2001, both assigned to the assignee of the present application and incorporated herein by reference.

[1156] In an embodiment, each MIMO terminal in the system estimates and sends back N_T post-processed SNR values for the N_T transmit antennas for each frequency subchannel group that may be separately assigned to the terminals. The SNRs from the active terminals may be evaluated by the scheduler to determine which terminal(s) to transmit data to and when, and the proper rate to use for each data stream transmitted to the selected terminals. MIMO terminals may be selected for data transmission based on a particular performance metric formulated to achieve the desired system goals. The performance metric may be based on one or more functions and any number of parameters. Various functions may be used to formulate the performance metric, such as the function of the achievable throughput for the MIMO terminals, which is shown above in equations (5) and (6).

[1157] In the N-SIMO mode, (up to) N_T independent data streams may be simultaneously transmitted by the base station from the N_T transmit antennas for each frequency subchannel group and targeted to (up to) N_T different SIMO terminals. To achieve high performance, the scheduler may consider a large number of possible terminal sets for data transmission. The scheduler then determines the best set of N_T SIMO terminals to transmit simultaneously for each frequency subchannel group. In a multiple-access communication system, there are generally constraints on satisfying certain requirements on a per terminal basis, such as maximum latency or average data rate. In this case, the scheduler can be designed to select the best set of terminals subject to these constraints.

[1158] In one implementation for the N-SIMO mode, the terminals use spatial equalization to process the receive signals, and the post-processed SNR corresponding to each data stream is provided to the base station. The scheduler then uses the information to select active terminals for data transmission and to assign transmission channels to the selected terminals.

[1159] In another implementation for the N-SIMO mode, the terminals use successive cancellation receiver processing to process the receive signal to achieve higher post-processed SNRs. With successive cancellation receiver processing, the post-processed SNRs for the transmitted data streams depend on the order in which the data streams are detected (i.e., demodulated and decoded). In some cases, a particular SIMO terminal may not be able to cancel the interference from a particular data stream designated for another terminal, since the coding and modulation scheme used for this data stream was selected based on the other terminal's post-processed SNR. For example, a transmitted data stream may be targeted for terminal u_x and coded and modulated for proper detection at a (e.g., 10 dB) post-processed SNR achieved at the target terminal u_x, but another terminal u_y may receive the same transmitted data stream at a worse post-processed SNR and is thus not able to properly detect the data stream. If the data stream intended for another terminal cannot be detected error free, then cancellation of the interference due to this data stream is not possible. Successive cancellation receiver processing is viable when the post-processed SNR corresponding to a transmitted data stream permits reliable detection.

[1160] The terminal can attempt to use successive cancellation receiver processing on all the other transmitted data streams not intended for it before attempting to process its own data stream to improve the reliability of the detection. However, in order for the system to capitalize on this improvement, the base station needs to know the hypothetical post-processed SNR given the interference from other antennas have been successfully cancelled. Independent constraints on the scheduler may result in a data rate assignment to these other antennas that precludes successful cancellation from being successful by the terminal. Thus there is no guarantee that the base station can select a data rate based on a post-processed SNR derived via successive cancellation receiver processing. However, the base station can use successive cancellation receiver processing on the uplink because it is the intended recipient of all data streams transmitted on the uplink.

[1161] In order for the scheduler to take advantage of the improvement in post-processed SNRs afforded by SIMO terminals using successive cancellation receiver processing, each such terminal can derive the post-processed SNRs corresponding to different possible orderings of detection for the transmitted data streams. The N_T transmitted data streams for each frequency subchannel group may be detected based on N_T factorial (i.e., $N_T!$) possible orderings at a SIMO terminal, and each ordering is associated with N_T post-processed SNR values. Thus, $N_T \cdot N_T!$ SNR values may be reported by each active terminal to the base station for each frequency subchannel group (e.g., if $N_T = 4$, then 96 SNR values may be reported by each SIMO terminal for each frequency subchannel group). The scheduler can then use the information to select terminals for data transmission and to further assign transmit antennas to the selected terminals.

[1162] If successive cancellation receiver processing is used at the terminals, then the scheduler can also consider the possible detection orderings for each terminal. However, a large number of these orderings are typically invalid because a particular terminal is able to properly detect data streams transmitted to other terminals due to the lower post-processed SNRs achieved at this terminal for the undetectable data streams.

[1163] In the mixed mode, the use of successive cancellation receiver processing by the (e.g., MIMO) terminals places additional constraints on the scheduler due to the dependencies introduced. These constraints may result in more hypothesized sets being evaluated, since in addition to considering different sets of terminals the scheduler needs to also consider the various orders for demodulating the data streams by each terminal in a given set. The assignment of the transmit antennas and the selection of the coding and modulation schemes would then take into account these dependencies in order to achieve high performance.

[1164] The set of transmit antennas at a base station may be a physically distinct set of "apertures", each of which may be used to directly transmit a respective data stream. Each aperture may be formed by a collection of one or more antenna elements that are distributed in space (e.g., physically located at a single site or distributed over multiple sites). Alternatively, the antenna apertures may be preceded by one or more (fixed) beam-forming matrices, with each matrix being used to synthesize a different set of antenna beams from the set of apertures. In this case, the above description for the transmit antennas applies analogously to the transformed antenna beams.

[1165] For the downlink, a number of fixed beam-forming matrices may be defined in advance, and the terminals may evaluate the post-processed SNRs for each of the possible matrices (or sets of antenna beams) and send SNR vectors back to the base station. Different performance (i.e., post-processed SNRs) is typically achieved for different sets of transformed antenna beams, and this is reflected in the reported SNR vectors. The base station may then perform scheduling and antenna assignment for each of the possible beam-forming matrices (using the reported SNR vectors), and select a particular beam-forming matrix as well as a set of terminals and their antenna assignments that achieve the best use of the available resources.

[1166] The use of beam-forming matrices affords additional flexibility in scheduling terminals and may further provide improved performance. As examples, the following situations may be well suited for beam-forming transformations:

- Correlation in the MIMO channel is high so that the best performance may be achieved with a small number of data streams. However, transmitting with only a subset of the available transmit antennas (and using only their associated transmit amplifiers) results in a smaller total transmit power. A transformation may be selected to use most or all of the transmit antennas (and their amplifiers) for the data streams to be sent. In this case, higher transmit power is achieved for the transmitted data streams.

- Physically dispersed terminals may be isolated somewhat by their locations. In this case, the terminals may be served by a standard FFT-type transformation of horizontally spaced apertures into a set of beams pointed at different azimuths.

Uplink Resource Allocation

[1167] On the uplink, since the base station is the intended recipient for the data transmissions from the scheduled terminals, the successive cancellation receiver processing technique may be used at the base station to process the transmissions from multiple terminals. This technique successively processes the N_R received signals a number of times to recover the signals transmitted from the terminals, with one transmitted signal being recovered for each iteration.

[1168] When using the successive cancellation receiver processing technique to process the received signals, the SNR associated with each received data stream is a function of the particular order in which the transmitted signals are processed at the base

station. The scheduling schemes can take this into account in selecting the best set of terminals for uplink data transmission.

[1169] FIG. 5 is a flow diagram of a process 500 to schedule terminals for uplink transmission. In this embodiment, the transmission channels are assigned to the active terminals by evaluating one frequency subchannel group at a time. The first frequency subchannel group is considered by setting the frequency index $k = 1$, at step 510. The best set of terminals for uplink transmission on the k-th frequency subchannel group is then determined starting at step 512.

[1170] Initially, one or more performance metrics to be used to select the best set of terminals for uplink transmission on the current frequency subchannel group are initialized, at step 512. Various performance metrics may be used, such as the performance metric that maximizes system throughput as described above. Also, terminal metrics such as post-processed SNRs for the signals transmitted from the terminals, the average throughput, and so on, may also be used in the evaluation.

[1171] A new set of one or more active terminals is then selected from among all active terminals desiring to transmit data in an upcoming time slot, at step 514. As noted above, the number of active terminals to be considered for scheduling may be limited (e.g., based on their priority). This set of selected terminals forms a hypothesis to be evaluated. For each selected terminal, the channel estimates for each transmit antenna to be used for uplink data transmission are retrieved, at step 516. For the MIMO mode, a single MIMO terminal is selected for evaluation for the k-th frequency subchannel group, and N_T vectors of channel estimates for N_T transmit antennas of this terminal are retrieved. For the N-SIMO mode, N_T SIMO terminals are selected for evaluation, and N_T channel estimate vectors for one transmit antenna at each of the N_T terminals are retrieved. And for the mixed mode, the N_T channel estimate vectors are retrieved for the combination of SIMO and MIMO terminals in the set. In any case, the N_T channel estimate vectors are used to form the channel response matrix $\underline{\mathbf{H}}(k)$ shown in equation (1), with each channel estimate vector corresponding to a column of the matrix $\underline{\mathbf{H}}(k)$. The set $\underline{\mathbf{u}}(k)$ identifies the terminals whose channel estimate vectors are included in the channel response matrix $\underline{\mathbf{H}}(k)$, where $\underline{\mathbf{u}}(k) = \{u_a(k), u_b(k), \ldots, u_{N_T}(k)\}$ and a MIMO terminal may be represented as multiple terminals in the set $\underline{\mathbf{u}}(k)$.

[1172] When the successive cancellation receiver processing technique is used at the base station, the order in which the terminals are processed directly impacts their performance. Thus, a particular new order is selected to process the terminals in the set $\underline{\mathbf{u}}(k)$, at step 518. This particular order forms a sub-hypothesis to be evaluated.

[1173] The sub-hypothesis is then evaluated and terminal metrics for the sub-hypothesis are provided, at step 520. The terminal metrics may be the post-processed SNRs for the signals (hypothetically) transmitted from the terminals in the set $\underline{\mathbf{u}}(k)$ to the base station. Step 520 may be achieved based on the successive cancellation receiver processing technique, which is described below in FIGS. 6A and 6B. The performance metric (e.g., the system throughput) corresponding to this sub-hypothesis is then determined (e.g., based on the post-processed SNRs for the terminals), at step 522. This performance metric is then used to update the performance metric for the best sub-hypothesis, also at step 522. Specifically, if the performance metric for the current sub-hypothesis is better than that for the best sub-hypothesis, then the current sub-hypothesis becomes the new best sub-hypothesis and the performance and terminal metrics corresponding to this sub-hypothesis are saved.

[1174] A determination is then made whether or not all sub-hypotheses for the current hypothesis have been evaluated, at step 524. If all sub-hypotheses have not been evaluated, then the process returns to step 518 and a different and not yet evaluated order for the terminals in the set $\underline{\mathbf{u}}(k)$ is selected for evaluation. Steps 518 through 524 are repeated for each sub-hypothesis to be evaluated.

[1175] If all sub-hypotheses for the current hypothesis have been evaluated, at step 524, then a determination is next made whether or not all hypotheses have been considered, at step 526. If all hypotheses have not been considered, then the process returns to step 514 and a different and not yet considered set of terminals is selected for evaluation. Steps 514 through 526 are repeated for each hypothesis to be considered.

[1176] If all hypotheses for the current frequency subchannel group have been evaluated, at step 526, then the results for the best sub-hypothesis for this frequency subchannel group are saved, at step 528. The best sub-hypothesis corresponds to a specific set of one or more active terminals that provides the best performance metric for the frequency subchannel group. If successive cancellation receiver processing is used at the base station, then the best sub-hypothesis is further associated with a specific

receiver processing order at the base station. The saved results may thus include the achievable SNRs for the terminals and the selected processing order.

[1177] If the scheduling scheme requires other system and terminal metrics to be maintained (e.g. the average throughput over the prior N_P time slots, latency for data transmission, and so on), then these metrics are updated for the current frequency subchannel group, at step 530. The terminal and system metrics may also be saved.

[1178] A determination is then made whether or not all frequency subchannel groups have been assigned for uplink transmission, at step 532. If all frequency subchannel groups have not been assigned, then the next frequency subchannel group is considered by incrementing the index k (i.e., $k = k + 1$), at step 534. The process then returns to step 512 to select the best set of terminals for uplink transmission on this new frequency subchannel group. Steps 512 through 534 are repeated for each frequency subchannel group to be assigned.

[1179] If all frequency subchannel groups have been assigned, at step 532, then the data rates and the coding and modulation schemes for the terminals in the best sub-hypotheses for each frequency subchannel group are determined (e.g., based on their SNRs), at step 536. A schedule indicative of the selected terminals and their assigned transmission channels and rates is formed and may be communicated to these terminals prior to the scheduled time slot, also at step 536. The uplink scheduling is typically performed for each scheduling interval.

[1180] FIG. 6A is a flow diagram for a successive cancellation receiver processing scheme 520a whereby the processing order is imposed by an ordered set of terminals. This flow diagram may be used for step 520 in FIG. 5. The processing shown in FIG. 6A is performed for a particular sub-hypothesis, which corresponds to a set of ordered terminals, $\underline{\mathbf{u}}(k) = \{u_a(k), u_b(k), \ldots, u_{N_T}(k)\}$. Initially, the first terminal in the ordered set is selected as the current terminal to be processed (i.e., $u_i = u_a(k)$), at step 612.

[1181] For the successive cancellation receiver processing technique, the base station first performs spatial or space-time equalization on the N_R received signals to attempt to separate the individual signals transmitted by the terminals in the set $\underline{\mathbf{u}}(k)$, at step 614. The spatial or space-time equalization may be performed as described below. The amount of achievable signal separation is dependent on the amount of correlation between the transmitted signals, and greater signal separation may be obtained if these signals are less correlated. Step 614 provides N_T post-processed signals derived from

the N_R received signals and corresponding to the N_T signals transmitted by the terminals in the set $\underline{\mathbf{u}}(k)$. As part of the spatial or space-time processing, the SNR corresponding to the post-processed signal for the current terminal u_i is also determined.

[1182] The post-processed signal for terminal u_i is further processed (i.e., "detected") to obtain a decoded data stream for the terminal, at step 616. The detection may include demodulating, deinterleaving, and decoding the post-processed signal to obtain the decoded data stream.

[1183] At step 618, a determination is made whether or not all terminals in the set $\underline{\mathbf{u}}(k)$ have been processed. If all terminals have been processed, then the SNRs of the terminals are provided, at step 626, and the receiver processing for this ordered set terminates. Otherwise, the interference due to the signal transmitted from terminal u_i on each of the received signals is estimated, at step 620. The interference may be estimated (e.g., as described below) based on the channel response matrix $\underline{\mathbf{H}}(k)$ for the terminals in the set $\underline{\mathbf{u}}(k)$. The estimated interference due to terminal u_i is then subtracted (i.e., canceled) from the received signals to derive modified signals, at step 622. These modified signals represent estimates of the received signals if terminal u_i had not transmitted (i.e., assuming that the interference cancellation was effectively performed). The modified signals are used in the next iteration to process the signal transmitted from the next terminal in the set $\underline{\mathbf{u}}(k)$. The next terminal in the set $\underline{\mathbf{u}}(k)$ is then selected as the (new) current terminal u_i, at step 624. In particular, $u_i = u_b(k)$ for the second iteration, $u_i = u_c(k)$ for the third iteration, and so on, and $u_i = u_{N_T}(k)$ for the last iteration for the ordered set $\underline{\mathbf{u}}(k) = \{u_a(k),\ u_b(k),\ ...,\ u_{N_T}(k)\}$.

[1184] The processing performed in steps 614 and 616 is repeated on the modified signals (instead of the received signals) for each subsequent terminal in the set $\underline{\mathbf{u}}(k)$. Steps 620 through 624 are also performed for each iteration except for the last iteration.

[1185] Using the successive cancellation receiver processing technique, for each hypothesis of N_T terminals, there are N_T factorial possible orderings (e.g., $N_T! = 24$ if $N_T = 4$). For each ordering of terminals within a particular hypothesis (i.e., for each sub-hypothesis), the successive cancellation receiver processing (step 520) provides a set of SNRs for the post-processed signals for these terminals, which may be expressed as:

$$\underline{\gamma}_{hyp,order}(k) = \{\gamma_1(k),\ \gamma_2(k),\ \ldots,\ \gamma_{N_T}(k)\}\ ,\qquad \text{Eq (18)}$$

where $\gamma_i(k)$ is the SNR for the k-th frequency subchannel group after the receiver processing at the i-th terminal in the sub-hypothesis.

[1186] Each sub-hypothesis is further associated with a performance metric, $R_{hyp,order}(k)$, which may be a function of various factors. For example, a performance metric based on the SNRs of the terminals may be expressed as shown in equation (4). In an embodiment, the performance metric for the sub-hypothesis is a function of the achievable throughputs for all N_T terminals in the set $\underline{\mathbf{u}}(k)$, which may be expressed as shown in equation (5), where the throughput $r_i(k)$ associated with the i-th terminal in the sub-hypothesis may be expressed as shown in equation (6).

[1187] The uplink scheduling scheme described in FIGS. 5 and 6A may be used to evaluate all possible orderings of each possible set of active terminals desiring to transmit data on the uplink. The total number of potential sub-hypotheses to be evaluated by the uplink scheduler can be quite large, even for a small number of active terminals. In fact, the total number of sub-hypotheses can be expressed as:

$$N_{sub-hyp} = N_G \cdot N_T!\begin{pmatrix} N_U \\ N_T \end{pmatrix} = \frac{N_G \cdot N_U!}{(N_U - N_T)!}\ ,\qquad \text{Eq (19)}$$

where N_U is the number of terminals to be considered for scheduling (again, a MIMO terminal may be represented as multiple terminals in the scheduling). For example, if $N_G = 16$, $N_U = 8$, and $N_T = 4$, then $N_{sub-hyp} = 26{,}880$. An exhaustive search may be used to determine the sub-hypothesis that provides the best system performance for each frequency subchannel group, as quantified by the performance metric used to select the best sub-hypothesis.

[1188] Similar to the downlink, a number of techniques may be used to reduce the complexity of the processing to schedule terminals for uplink transmission. Some scheduling schemes based on some of these techniques are described below. Other scheduling schemes may also be implemented and are within the scope of the invention. These scheduling schemes may also provide high system performance while reducing the amount of processing required to schedule terminals for uplink data transmission.

[1189] In a second uplink scheduling scheme, the terminals included in each hypothesis are processed in a specific order that is determined based on a particular defined rule. In an embodiment, this scheme relies on the successive cancellation receiver processing to determine the specific order for processing the terminals in the hypothesis. For example and as described below, for each iteration, the successive cancellation receiver processing scheme can recover the transmitted signal having the best SNR after equalization. In this case, the processing order is determined based on the post-processed SNRs for the terminals in the hypothesis.

[1190] FIG. 6B is a flow diagram for a successive cancellation receiver processing scheme 520b whereby the processing order is determined based on the post-processed SNRs. This flow diagram may also be used for step 520 in FIG. 5. However, since the processing order is determined based on the post-processed SNRs achieved by the successive cancellation receiver processing, only one sub-hypothesis is effectively evaluated for each hypothesis, and steps 518 and 524 in FIG. 5 may be omitted.

[1191] Initially, spatial or space-time equalization is performed on the received signals to attempt to separate the individual transmitted signals, at step 614. The SNRs of the transmitted signals after the equalization are then estimated, at step 615. In an embodiment, the transmitted signal corresponding to the terminal with the best SNR is selected and further processed (i.e., demodulated and decoded) to obtain a corresponding decoded data stream, at step 616. At step 618, a determination is made whether or not all transmitted signals (i.e., all terminals in the hypothesis) have been processed. If all terminals have been processed, then the processing order of the terminals and their SNRs are provided, at step 628, and the receiver processing for this terminal set terminates. Otherwise, the interference due to the transmitted signal just processed is estimated, at step 620, and subtracted (i.e., canceled) from the received signals to derive the modified signals, at step 622. Steps 614, 616, 618, 620, and 622 in FIG. 6B correspond to identically numbered steps in FIG. 6A.

[1192] In a third uplink scheduling scheme, the terminals included in each hypothesis are processed based on a specific order. With successive cancellation receiver processing, the SNR of an unprocessed terminal improves with each iteration, as the interference from each processed terminal is removed. Thus, on average, the first terminal to be processed will have the lowest SNR, the second terminal to be processed will have the second to lowest SNR, and so on. Using this knowledge, the processing

order for the terminals may be specified for a hypothesis. The processing order represents another degree of freedom that may be used by the scheduler to achieve the system goals and requirements.

[1193] In one embodiment of the third uplink scheduling scheme, the processing order for each hypothesis is selected based on the priority of the terminals in the hypothesis. For example, the lowest priority terminal in the hypothesis may be processed first, the next lowest priority terminal may be processed next, and so on, and the highest priority terminal may be processed last. This embodiment allows the highest priority terminal to achieve the highest SNR possible for the hypothesis, which in turn supports the highest possible data rate. In this manner, the terminals may be assigned transmission channels in a particular order, based on their priority, such that the highest priority terminal is assigned the highest possible data rate. In another embodiment of the third uplink scheduling scheme, the processing order for each hypothesis is selected based on user payload, latency requirements, emergency service priority, and so on.

[1194] In a fourth uplink scheduling scheme, the terminals are scheduled based on their priority, which may be determined based on one or more metrics (e.g., average throughput), system constraints and requirements (e.g., maximum latency), other factors, or a combination thereof, as described above. For each scheduling interval, a number of highest priority terminals may be considered for scheduling.

[1195] FIG. 7 is a flow diagram for a priority-based uplink scheduling scheme 700 whereby a set of N_T highest priority terminals is considered for scheduling for each frequency subchannel group. Initially, the first frequency subchannel group is considered by setting the frequency index $k = 1$, at step 710. The spatial subchannels for the k-th frequency subchannel group are then assigned to the terminals for uplink transmission starting at step 712.

[1196] The scheduler examines the priority for all active terminals in the list and selects the set of N_T highest priority terminals, at step 712. The remaining active terminals in the list are not considered for scheduling for this frequency subchannel group in this scheduling interval. The channel estimates for each selected terminal are retrieved and used to form the channel response matrix $\underline{\mathbf{H}}(k)$, at step 714.

[1197] Each sub-hypothesis of the hypothesis formed by the N_T selected terminals is then evaluated, and the corresponding vector of post-processed SNRs, $\underline{\gamma}_{hyp,order}(k)$, for each sub-hypothesis is derived, at step 716. The best sub-hypothesis is selected, and the

data rates and the coding and modulation schemes for the terminals in the best sub-hypothesis are determined (e.g., based on their achieved SNRs), at step 718. The metrics of the active terminals in the list and the system metrics are then updated, at step 720.

[1198] A determination is then made whether or not all frequency subchannels have been assigned for uplink transmission, at step 722. If all frequency subchannels have not been assigned, then the next frequency subchannel group is considered by incrementing the index k (i.e., $k = k + 1$), at step 724. The process then returns to step 712 to assign the spatial subchannels of this new frequency subchannel group to the same or a different set of terminals. Steps 712 through 724 are repeated for each frequency subchannel group to be assigned.

[1199] If all frequency subchannel groups have been assigned, at step 722, then a schedule indicative of the selected terminals and their assigned transmission channels and rates may be formed and communicated to these terminals, at step 726. The process then terminates for this scheduling interval.

[1200] The uplink scheduling of terminals based on priority is also described in U.S. Patent Application Serial No. 09/859,346, entitled "Method and Apparatus for Allocating Uplink Resources in a Multiple-Input Multiple-Output (MIMO) Communication System," filed May 16, 2001, and U.S. Patent No. 5,923,650, entitled "Method and Apparatus for Reverse Link Rate Scheduling," issued July 13, 1999. These patent and patent application are assigned to the assignee of the present application and incorporated herein by reference.

[1201] The same target setpoint may be used for all data streams received at the base station. However, this common setpoint for all received data streams is not a requirement. Other uplink scheduling schemes that select terminals for data transmission, assign transmission channels to the selected terminals, and further select setpoints to be used for the assigned transmission channels may also be devised. A particular setpoint may be achieved for a data stream via a power control mechanism that direct the terminal to adjust its transmit power for the data stream such that the received SNR for the data stream is approximately equal to the setpoint.

[1202] Various uplink scheduling schemes may be devised with non-uniform setpoints for the data streams transmitted from the scheduled terminals. In one embodiment, higher setpoints may be used for higher priority terminals, and lower

setpoints may be used for lower priority terminals. In another embodiment, the setpoints may be selected such that a desired data rate is achieved for each of the scheduled terminals. In yet another embodiment, the setpoints may be selected to achieve high system throughput, which may be possible by using higher setpoints for better transmission channels and lower setpoints for poor transmission channels. Other schemes to select different setpoints for different transmission channels to achieve the desired results may also be implemented, and this is within the scope of the invention.

[1203] Similar to the downlink, it is also not necessary to use all of the available transmission channels for uplink data transmission. In one embodiment, only transmission channels with achieved SNRs above a particular threshold SNR are selected for use, and transmission channels with achieved SNRs below this threshold SNR are not used.

[1204] For many of the uplink scheduling schemes described above, the successive cancellation receiver processing technique is used to process the received signals at the base station, which may provide improved SNRs and thus higher throughput. However, the uplink scheduling may also be performed without the use of successive cancellation receiver processing at the base station. For example, the base station may simply use spatial or space-time equalization to process the received signals to recover the transmitted signals. It can be shown that substantial gains may be achieved by exploiting the multi-user diversity environment and/or the frequency signatures of the terminals in scheduling uplink data transmission (i.e., without relying on successive cancellation receiver processing at the base station).

[1205] Other uplink scheduling schemes may also be implemented, and this is within the scope of the invention. For a FDM-TDM uplink scheduling scheme, one MIMO terminal may be assigned all of the spatial subchannels for each frequency subchannel group, and the frequency signatures of the terminals may be considered in the uplink scheduling to achieve high performance. For a SDMA-TDM uplink scheduling scheme, all frequency subchannels of each spatial subchannel may be assigned to a single terminal, which may be a SIMO or MIMO terminal.

Other Scheduling Considerations

[1206] For both the downlink and uplink, if partial-CSI (e.g., the post-processed SNR) is used to schedule terminals for data transmission, then a common coding and

modulation scheme may be used for all transmission channels assigned to a given terminal, or a different coding and modulation scheme may be used for each assigned transmission channel. The use of a common coding and modulation scheme for all assigned transmission channels can simplify the processing at both the terminal and the base station. The scheduler may be designed to take this into consideration when scheduling terminals for data transmission on the available transmission channels. For example, it may be preferable to assign transmission channels having similar transmission capacities (e.g., similar SNRs) to the same terminal so that a common coding and modulation scheme may be used for the data transmission on the multiple transmission channels assigned to this terminal.

[1207] For both the downlink and uplink, the scheduling schemes can be designed to consider sets of terminals that have similar link margins. Terminals may be grouped according to their link margin properties. The scheduler may then consider combinations of terminals in the same "link margin" group when searching for mutually compatible spatial signatures. The grouping of terminals according to link margin may improve the overall spectral efficiency of the scheduling schemes compared to that achieved by ignoring link margins. Moreover, by scheduling terminals with similar link margins to transmit concurrently, power control may be more easily exercised (e.g., on the entire set of terminals) to improve overall spectral reuse. This may be viewed as a combination of adaptive reuse scheduling in combination with SDMA for SIMO/MIMO (which relies on spatial processing at the receiver to separate the multiple transmitted data streams) or MISO (which relies on beam-steering by the transmitter to separate the multiple transmitted data streams). Moreover, a scheduling scheme that evaluates the hybrid of these two (beams and margins) may also be implement, and this is within the scope of the invention.

[1208] Scheduling based on link margins and adaptive reuse are described in further detail in U.S. Patent Application Serial No. 09/532,492, entitled "High Efficiency, High Performance Communications System Employing Multi-Carrier Modulation," filed March 30, 2000, and U.S. Patent Application Serial No. 09/848,937, entitled "Method and Apparatus for Controlling Uplink Transmissions of a Wireless Communication System," filed May 3, 2001, both assigned to the assignee of the present application and incorporated herein by reference.

[1209] For simplicity, various scheduling schemes have been described whereby (1) a set of N_T terminals is selected for downlink or uplink transmission for a given frequency subchannel group (where a MIMO terminal may represent multiple ones of these N_T terminals), with each terminal being assigned to one spatial subchannel, (2) the number of transmit antennas is equal to the number of receive antennas (i.e., $N_T = N_R$), and (3) one independent data stream is transmitted on each spatial subchannel of each frequency subchannel group. In this case, the number of data streams for each frequency subchannel group is equal to the number of spatial subchannels, and each of the N_T terminals in the set is effectively assigned to a respective spatial subchannel.

[1210] For the downlink, each scheduled terminal may be equipped with more receive antennas than the total number of data streams. Moreover, multiple scheduled terminals may share a particular transmit antenna at the base station. The sharing may be achieved via time division multiplexing (e.g., assigning different fractions of a time slot to different terminals), frequency division multiplexing (e.g., assigning different frequency subchannels in each frequency subchannel group to different terminals), code division multiplexing (e.g., assigning different orthogonal codes to different terminals), some other multiplexing schemes, or any combination of the multiplexing schemes.

[1211] For the uplink, the scheduled terminals may also share a multiplexed array of receive antennas at the base station. In this case, the total number of transmit antennas for the scheduled terminals may be greater than the number of receive antennas at the base station, and the terminals may share the available transmission channels using another multiple-access technique (e.g., time, frequency, and/or code division multiplexing).

[1212] The scheduling schemes described herein select terminals and assign transmission channels to the selected terminals based on channel state information, which may comprise post-processed SNRs. The post-processed SNRs for the terminals are dependent on the particular transmit power level used for the data streams. For simplicity, the same transmit power level is assumed for all data streams (i.e., no power control of the transmit power).

[1213] However, by allocating different amounts of transmit power to different data streams and/or by controlling the transmit power for each data stream, the achievable SNRs may be adjusted. For the downlink, by decreasing the transmit power for a particular data stream via power control, the SNR associated with that data stream is

reduced, the interference caused by this data stream on other data streams would also be reduced, and other data streams may be able to achieve better SNRs. For the uplink, by decreasing the transmit power of a particular terminal via power control, the SNR for this terminal is reduced, the interference due to this terminal would also be reduced, and other terminals may be able to achieve better SNRs. Power control of (and power allocation among) multiple terminals simultaneously sharing non-orthogonal spatial channels may be achieved by placing various constraints to ensure system stability, as described above. Thus, transmit power allocation and/or power control may also be used in conjunction with the scheduling schemes described herein, and this is within the scope of the invention.

[1214] The downlink and uplink scheduling schemes described herein may be designed to support a number of features. First, the scheduling schemes can support mixed mode operation whereby any combination of SIMO and MIMO terminals may be scheduled for data transmission over a "channel", which may be a time slot, a frequency band, a code channel, and so on. Second, the scheduling schemes provide a schedule for each scheduling interval that includes a set of "mutually compatible" terminals based on their spatial and frequency signatures. Mutual compatibility may be taken to mean co-existence of transmission on the same channel and at the same time given specific constraints regarding terminals' data rate requirements, transmit power, link margin, capability between SIMO and MIMO terminals, and possibly other factors. Third, the scheduling schemes support variable data rate adaptation based on the SNRs of the post-processed signals for the terminals. Each scheduled terminal is informed when to communicate, which data rate(s) to use (e.g., on a per data stream basis), and the particular mode (e.g., SIMO, MIMO).

MIMO-OFDM System

[1215] FIG. 8A is a block diagram of a base station 104 and two terminals 106 within MIMO-OFDM system 100 for downlink data transmission. At base station 104, a data source 808 provides data (i.e., information bits) to a transmit (TX) data processor 810. For each independent data stream, TX data processor 810 (1) codes the data based on a particular coding scheme, (2) interleaves (i.e., reorders) the coded bits based on a particular interleaving scheme, and (3) maps the interleaved bits into modulation symbols for one or more transmission channels selected for use for that data stream.

The coding increases the reliability of the data transmission. The interleaving provides time diversity for the coded bits, permits the data to be transmitted based on an average SNR for the transmission channels, combats fading, removes correlation between coded bits used to form each modulation symbol, and may further provide frequency diversity if the coded bits are transmitted over multiple frequency subchannels. The coding and modulation (i.e., symbol mapping) may be performed based on control signals provided by a controller 830.

[1216] A TX MIMO processor 820 receives and demultiplexes the modulation symbols from TX data processor 810 and provides a stream of symbol vectors for each transmit antenna used for data transmission, one symbol vector per symbol period. Each symbol vector includes up to N_F modulation symbols for the N_F frequency subchannels of the transmit antenna. TX MIMO processor 820 may further precondition the modulation symbols if full CSI processing is performed (e.g., if the channel response matrix $\underline{\mathbf{H}}(k)$ is available). MIMO and full-CSI processing is described in further detail in the aforementioned U.S. Patent Application Serial No. 09/993,087. Each symbol vector stream is then received and modulated by a respective modulator (MOD) 822 and transmitted via an associated antenna 824.

[1217] At each terminal 106 to which a data transmission is directed, antennas 852 receive the transmitted signals, and the received signal from each antenna is provided to a respective demodulator (DEMOD) 854. Each demodulator (or front-end unit) 854 performs processing complementary to that performed at modulator 822. The received modulation symbols from all demodulators 854 are then provided to a receive (RX) MIMO/data processor 860 and processed to recover one or more data streams transmitted to the terminal. RX MIMO/data processor 860 performs processing complementary to that performed by TX data processor 810 and TX MIMO processor 820 and provides decoded data to a data sink 862. The processing by terminal 106 is described in further detail below.

[1218] At each active terminal 106, RX MIMO/data processor 860 further estimates the channel conditions for the downlink and provides channel state information (CSI) indicative of the estimated channel conditions. The CSI may comprise post-processed SNRs, channel gain estimates, and so on. A controller 870 receives and may further transform the downlink CSI (DL CSI) into some other form (e.g., rate). The downlink CSI is processed (e.g., coded and symbol mapped) by a TX data processor 880, further

processed by a TX MIMO processor 882, modulated by one or more modulators 854, and transmitted back to base station 104 via an uplink (or feedback) channel. The downlink CSI may be reported by the terminal using various signaling techniques, as described below.

[1219] At base station 104, the transmitted feedback signal is received by antennas 824, demodulated by demodulators 822, and processed by a RX MIMO/data processor 840 in a complementary manner to that performed by TX data processor 880 and TX MIMO processor 882. The reported downlink CSI is then provided to controller 830 and a scheduler 834.

[1220] Scheduler 834 uses the reported downlink CSI to perform a number of functions such as (1) selecting the best set of terminals for downlink data transmission and (2) assigning the available transmission channels to the selected terminals. Scheduler 834 or controller 830 may further use the reported downlink CSI to determine the coding and modulation scheme to be used for each data stream. Scheduler 834 may schedule terminals to achieve high throughput and/or based on some other performance criteria or metrics.

[1221] FIG. 8B is a block diagram of a base station 104 and two terminals 106 for uplink data transmission. At each terminal scheduled for data transmission on the uplink, a data source 878 provides data to TX data processor 880, which codes, interleaves, and maps the data into modulation symbols. If multiple transmit antennas are used for uplink data transmission, TX MIMO processor 882 receives and further processes the modulation symbols to provide a stream of modulation symbol vectors for each antenna used for data transmission. Each symbol vector stream is then received and modulated by a respective modulator 854 and transmitted via an associated antenna 852.

[1222] At base station 104, antennas 824 receive the transmitted signals, and the received signal from each antenna is provided to a respective demodulator 822. Each demodulator 822 performs processing complementary to that performed at modulator 854. The modulation symbols from all demodulators 822 are then provided to RX MIMO/data processor 840 and processed to recover the data streams transmitted by the scheduled terminals. RX MIMO/data processor 840 performs processing complementary to that performed by TX data processor 880 and TX MIMO processor 882 and provides decoded data to a data sink 842.

[1223] For each terminal 106 desiring to transmit data on the uplink during an upcoming scheduling interval (or only the N_T or N_X highest priority terminals), RX MIMO/data processor 840 further estimates the channel conditions for the uplink and derives uplink CSI (UL CSI), which is provided to controller 830. Scheduler 834 may also receive and use the uplink CSI to perform a number of functions such as (1) selecting the best set of terminals for data transmission on the uplink, (2) determining a particular processing order for the data streams from the selected terminals, and (3) determining the rate to be used for each data stream. For each scheduling interval, scheduler 834 provides an uplink schedule that indicates which terminal(s) have been selected for data transmission and their assigned transmission channels and rates. The rate for each data stream may include the date rate and coding and modulation scheme to be used for the data stream.

[1224] TX data processor 810 receives and processes the uplink schedule, and provides processed data indicative of the schedule to one or more modulators 822. Modulator(s) 822 further condition the processed data and transmit the uplink schedule to the terminals via the wireless link. The uplink schedule may be sent to the terminal using various signaling and messaging techniques.

[1225] At each active terminal 106, the transmitted signals are received by antennas 852, demodulated by demodulators 854, and provided to RX MIMO/data processor 860. Processor 860 performs processing complementary to that performed by TX MIMO processor 820 and TX data processor 810 and recovers the uplink schedule for that terminal (if any), which is then provided to controller 870 and used to control the uplink transmission by the terminal.

[1226] In FIGS. 8A and 8B, scheduler 834 is shown as being implemented within base station 104. In other implementations, scheduler 834 may be implemented within some other element of MIMO-OFDM system 100 (e.g., a base station controller that couples to and interacts with a number of base stations).

[1227] FIG. 9 is a block diagram of an embodiment of a transmitter unit 900. For clarity, transmitter unit 900 is described as being the transmitter portion of base station 104 in FIGS. 8A and 8B. However, transmitter unit 900 may also be used for the transmitter portion of each terminal for uplink transmissions.

[1228] Transmitter unit 900 is capable of processing multiple data streams for one or more terminals based on the available CSI (e.g., as reported by the terminals).

Transmitter unit 900 includes (1) a TX data processor 814x that receives and processes information bits to provide modulation symbols and (2) a TX MIMO processor 820x that demultiplexes the modulation symbols for the N_T transmit antennas.

[1229] In the specific embodiment shown in FIG. 9, TX data processor 814x includes a demultiplexer 908 coupled to a number of channel data processors 910, one processor for each of N_D independent data streams to be transmitted to the terminal(s). Demultiplexer 908 receives and demultiplexes the aggregate information bits into N_D data streams, each of which may be transmitted over one or more transmission channels. Each data stream is provided to a respective channel data processor 910.

[1230] In the embodiment shown in FIG. 9, each channel data processor 910 includes an encoder 912, a channel interleaver 914, and a symbol mapping element 916. Encoder 912 codes the information bits in the received data stream based on a particular coding scheme to provide coded bits. Channel interleaver 914 interleaves the coded bits based on a particular interleaving scheme to provide diversity. And symbol mapping element 916 maps the interleaved bits into modulation symbols for the one or more transmission channels used for transmitting the data stream.

[1231] Pilot data (e.g., data of known pattern) may also be coded and multiplexed with the processed information bits. The processed pilot data may be transmitted (e.g., in a time division multiplexed (TDM) or code division multiplexed (CDM) manner) in all or a subset of the transmission channels used to transmit the information bits. The pilot data may be used at the receiver systems to perform channel estimation.

[1232] As shown in FIG. 9, the data coding, interleaving, and modulation (or a combination thereof) may be adjusted based on the available CSI (e.g., as reported by the receiver systems). In one coding and modulation scheme, adaptive coding is achieved by using a fixed base code (e.g., a rate 1/3 Turbo code) and adjusting the puncturing to achieve the desired code rate, as supported by the SNRs of the transmission channels used to transmit the data. For this scheme, the puncturing may be performed after the channel interleaving. In another coding and modulation scheme, different coding schemes may be used based on the reported CSI. For example, each of the data streams may be coded with an independent code. With this scheme, the successive cancellation receiver processing technique may be used at the receivers to detect and decode the data streams to derive a more reliable estimate of the transmitted data streams, as described in further detail below.

[1233] Symbol mapping element 916 can be designed to group sets of interleaved bits to form non-binary symbols, and to map each non-binary symbol to a point in a signal constellation corresponding to a particular modulation scheme (e.g., QPSK, M-PSK, M-QAM, or some other scheme) selected for the data stream. Each mapped signal point corresponds to a modulation symbol. The number of information bits that may be transmitted for each modulation symbol for a particular level of performance (e.g., one percent PER) is dependent on the SNRs of the transmission channels used to transmit the data stream. Thus, the coding and modulation scheme for each data stream may be selected based on the available CSI. The channel interleaving may also be adjusted based on the available CSI.

[1234] The modulation symbols from TX data processor 814x are provided to TX MIMO processor 820x. TX MIMO processor 820x receives N_D modulation symbol streams from N_D channel data processors 910 and demultiplexes the received modulation symbols into N_T symbol vector streams, V_1 through V_{Nt}, one symbol vector stream for each antenna used to transmit data. Each symbol vector stream is provided to a respective modulator 822. In the embodiment shown in FIG. 9, each modulator 822 includes an inverse fast Fourier transform (IFFT) processor 940, a cyclic prefix generator 942, and a transmitter (TMTR) 944.

[1235] IFFT processor 940 converts each received symbol vector into its time-domain representation (which is referred to as an OFDM symbol) using the IFFT. IFFT processor 940 can be designed to perform the IFFT on any number of frequency subchannels (e.g., 8, 16, 32, ... , N_F, ...). In an embodiment, for each symbol vector converted to an OFDM symbol, cyclic prefix generator 942 repeats a portion of the time-domain representation of the OFDM symbol to form a "transmission symbol" for a specific transmit antenna. The cyclic prefix ensures that the transmission symbol retains its orthogonal properties in the presence of multipath delay spread, thereby improving performance against deleterious path effects. The implementation of IFFT processor 940 and cyclic prefix generator 942 is known in the art and not described in detail herein.

[1236] Transmitter 944 then converts the time-domain transmission symbols from an associated cyclic prefix generator 942 into an analog signal, and further amplifies, filters, quadrature modulates, and upconverts the analog signal to provide a modulated

signal suitable for transmission over the wireless link. The modulated signals from transmitters 944 are then transmitted from antennas 824 to the terminals.

[1237] An example MIMO-OFDM system is described in the aforementioned U.S. Patent Application Serial No. 09/532,492. OFDM modulation is also described in a paper entitled "Multicarrier Modulation for Data Transmission : An Idea Whose Time Has Come," by John A.C. Bingham, IEEE Communications Magazine, May 1990, which is incorporated herein by reference.

[1238] FIG. 9 shows an example coding and modulation scheme that may be used with full or partial CSI to provide improved performance (e.g., high throughput). Some other coding and modulation schemes are described in further detail in the aforementioned U.S. Patent Application Serial Nos. 09/854,235, 09/826,481, and 09/956,449, and in U.S Patent Application Serial No. 09/776,075, entitled "Coding Scheme for a Wireless Communication System," filed February 1, 2001, which is assigned to the assignee of the present application and incorporated herein by reference. Still other coding and modulation schemes may also be used, and this is within the scope of the invention.

[1239] FIG. 10A is a block diagram of an embodiment of a receiver unit 1000a. For clarity, receiver unit 1000a is described as being the receiver portion of one terminal 106 in FIGS. 8A and 8B. However, receiver unit 1000a may also be used for the receiver portion of base station 104 for uplink transmissions.

[1240] The transmitted signals from N_T transmit antennas are received by each of N_R antennas 852a through 852r, and the received signal from each antenna is routed to a respective demodulator 854 (which is also referred to as a front-end processor). Each demodulator 854 conditions (e.g., filters and amplifies) a respective received signal, downconverts the conditioned signal to an intermediate frequency or baseband, and digitizes the downconverted signal to provide data samples. Each demodulator 854 may further demodulate the data samples with a recovered pilot.

[1241] Each demodulator 854 also performs processing complementary to that performed by modulator 822 shown in FIG. 9. For OFDM, each demodulator 854 includes an FFT processor and a demultiplexer (both of which are not shown in FIG. 10A for simplicity). The FFT processor generates transformed representations of the data samples and provides a stream of symbol vectors. Each symbol vector includes N_F symbols received for N_F frequency subchannels, and one vector is provided for each

symbol period. The N_R symbol vector streams from the FFT processors of all N_R demodulators are then provided to the demultiplexer, which demultiplexes each symbol vector stream into N_G received symbol vector streams for the N_G frequency subchannel groups. Each received symbol vector includes N_k received symbols for the N_k frequency subchannels in the k-th frequency subchannel group, where $1 \leq N_k \leq N_F$. The demultiplexer may then provide up to $N_G \cdot N_R$ received symbol vector streams for the N_G frequency subchannel groups in the N_R received signals.

[1242] Within a RX MIMO/data processor 860a, a spatial/space-time processor 1010 is used to perform MIMO processing for the received symbols for each frequency subchannel group used for data transmission. One spatial/space-time processor may be used to perform the MIMO processing for each frequency subchannel group, or one spatial/space-time processor may be used to perform the MIMO processing for all frequency subchannel groups (e.g., in a time division multiplexed manner).

[1243] Spatial/space-time processor 1010 may be designed to perform spatial processing or space-time processing on the received symbols to provide estimates of the transmitted modulation symbols. Spatial processing may be used for a non-dispersive channel (i.e., a flat fading channel) to null out the undesired signals and/or to maximize the received SNR of each of the constituent signals in the presence of noise and interference from the other signals. The spatial processing may be performed based on a channel correlation matrix inversion (CCMI) technique, a minimum mean square error (MMSE) technique, a full-CSI technique, or some other technique. Space-time processing may be used for a dispersive channel (i.e., a frequency selective fading channel) to ameliorate both "crosstalk" from the other transmitted signals as well as inter-symbol interference (ISI) from all of the transmitted signals due to dispersion in the channel. The space-time processing may be performed based on a MMSE linear equalizer (MMSE-LE), a decision feedback equalizer (DFE), a maximum-likelihood sequence estimator (MLSE), or some other technique. Spatial and space-time processing is described in further detail in the aforementioned U.S. Patent Application Serial No. 09/993,087.

[1244] For a particular frequency subchannel group, spatial/space-time processor 1010 receives and processes N_R received symbol vector streams and provides N_T recovered symbol vector streams. Each recovered symbol vector includes up to N_k recovered symbols that are estimates of the N_k modulation symbols transmitted on the

N_k frequency subchannels of the k-th frequency subchannel group in one symbol period. Spatial/space-time processor 1010 may further estimate the post-processed SNR for each received data stream. The SNR estimate may be derived as described in the aforementioned U.S. Patent Application Serial Nos. 09/956,449, 09/854,235, and 09/993,087.

[1245] A selector 1012 receives the N_T recovered symbol vector streams from spatial/space-time processor 1010 and extracts the recovered symbols corresponding to the one or more data streams to be recovered. Alternatively, the desired recovered symbols are extracted within spatial/space-time processor 1010. In any case, the desired recovered symbols are extracted and provided to a RX data processor 1020.

[1246] Within RX data processor 1020, a demodulation element 1022 demodulates each recovered symbol in accordance with a demodulation scheme (e.g., M-PSK, M-QAM) used for that symbol at the transmitter unit. The demodulated data is then de-interleaved by a de-interleaver 1024 and the de-interleaved data is further decoded by a decoder 1026. The demodulation, deinterleaving, and decoding are performed in a complementary manner to the modulation, interleaving, and coding performed at the transmitter unit. For example, a Turbo decoder or a Viterbi decoder may be used for decoder 1026 if Turbo or convolutional coding, respectively, is performed at the transmitter unit. The decoded data stream from decoder 1026 represents an estimate of the transmitted data stream.

[1247] FIG. 10B is a block diagram of a receiver unit 1000b capable of implementing the successive cancellation receiver processing technique. Receiver unit 1000b may also be used for the receiver portion of base station 104 or terminal 106. The transmitted signals are received by each of N_R antennas 852, and the received signal from each antenna is routed to a respective demodulator 854. Each demodulator 854 processes a respective received signal and provides a stream of received symbols to a RX MIMO/data processor 860b. RX MIMO/data processor 860b may be used to process the N_R received symbol vector streams from the N_R receive antennas for each frequency subchannel group used for data transmission, where each received symbol vector includes N_k received symbols for the N_k frequency subchannels in the k-th frequency subchannel group.

[1248] In the embodiment shown in FIG. 10B, RX MIMO/data processor 860b includes a number of successive (i.e., cascaded) receiver processing stages 1050, one

stage for each of the transmitted signals to be recovered. In one transmit processing scheme, one independent data stream is transmitted on each spatial subchannel of each frequency subchannel group. For this transmit processing scheme, the number of data streams for each frequency subchannel group is equal to the number of transmitted signals, which is also equal to the number of transmit antennas used for data transmission (which may be all or a subset of the available transmit antennas). For clarity, RX MIMO/data processor 860b is described for this transmit processing scheme.

[1249] Each receiver processing stage 1050 (except for the last stage 1050n) includes a channel MIMO/data processor 1060 coupled to an interference canceller 1070, and the last stage 1050n includes only channel MIMO/data processor 1060n. For the first receiver processing stage 1050a, channel MIMO/data processor 1060a receives and processes the N_R received symbol vector streams from demodulators 854a through 854r to provide a decoded data stream for the first transmitted signal. And for each of the second through last stages 1050b through 1050n, the channel MIMO/data processor 1060 for that stage receives and processes the N_R modified symbol vector streams from the interference canceller 1070 in the preceding stage to derive a decoded data stream for the transmitted signal being recovered by that stage. Each channel MIMO/data processor 1060 further provides CSI (e.g., the SNR) for the associated transmission channel.

[1250] For the first receiver processing stage 1050a, interference canceller 1070a receives the N_R received symbol vector streams from all N_R demodulators 854. And for each of the second through second-to-last stages, interference canceller 1070 receives the N_R modified symbol vector streams from the interference canceller in the preceding stage. Each interference canceller 1070 also receives the decoded data stream from the channel MIMO/data processor 1060 within the same stage, and performs the processing (e.g., coding, interleaving, and modulation) to derive N_T remodulated symbol vector streams that are estimates of the N_T transmitted modulation symbol vector streams for the frequency subchannel group.

[1251] The N_T remodulated symbol vector streams (for the *n*-th iteration) are further processed with the estimated channel response to provide estimates, $\hat{\underline{\mathbf{i}}}^n$, of the interference due to the decoded data stream. The estimates $\hat{\underline{\mathbf{i}}}^n$ include N_R vectors, with each vector being an estimate of a component in one of the N_R received signals due to

the decoded data stream. These components are interference to the remaining (not yet detected) transmitted signals included in the N_R received signals. Thus, the interference estimates, $\hat{\underline{\mathbf{i}}}^{n}$, are subtracted (i.e., canceled) from the received symbol vector streams, $\underline{\mathbf{r}}^{n}$, to provide N_R modified symbol vector streams, $\underline{\mathbf{r}}^{n+1}$, having the components from the decoded data stream removed. The modified symbol vector streams, $\underline{\mathbf{r}}^{n+1}$, are provided to the next receiver processing stage, as shown in FIG. 10B. Each interference canceller 1070 thus provides N_R modified symbol vector streams that include all but the cancelled interference components. Controller 870 may be used to direct various steps in the successive cancellation receiver processing.

[1252] The successive cancellation receiver processing technique is described in further detail in the aforementioned U.S Patent Application Serial Nos. 09/854,235 and 09/993,087, and by P.W. Wolniansky *et al.* in a paper entitled "V-BLAST: An Architecture for Achieving Very High Data Rates over the Rich-Scattering Wireless Channel," Proc. ISSSE-98, Pisa, Italy, which is incorporated herein by reference.

[1253] FIG. 10B shows a receiver structure that may be used in a straightforward manner when one independent data stream is transmitted over each transmit antenna of each frequency subchannel group. In this case, each receiver processing stage 1050 may be operated to recover one of the transmitted data streams and to provide the decoded data stream corresponding to the recovered data stream.

[1254] For some other transmit processing schemes, a data stream may be transmitted over multiple transmit antennas, frequency subchannels, and/or time intervals to provide spatial, frequency, and/or time diversity, respectively. For these schemes, the receiver processing initially derives a received symbol stream for each transmit antenna of each frequency subchannel. Modulation symbols for multiple transmit antennas, frequency subchannels, and/or time intervals may then be combined in a complementary manner as the demultiplexing performed at the transmitter unit. The stream of combined symbols is then processed to recover the transmitted data stream.

[1255] For simplicity, the receiver architecture shown in FIG. 10B provides the (received or modified) symbol vector streams to each receiver processing stage 1050, and these streams have the interference components due to previously decoded data streams removed (i.e., canceled). In the embodiment shown in FIG. 10B, each stage removes the interference components due to the data stream decoded by that stage. In

some other designs, the received symbol vector streams may be provided to all stages, and each stage may perform the cancellation of interference components from all previously decoded data streams (which may be provided from preceding stages). The interference cancellation may also be skipped for one or more stages (e.g., if the SNR for the data stream is high). Various modifications to the receiver architecture shown in FIG. 10B may be made and are within the scope of the invention.

[1256] FIGS. 10A and 10B represent two embodiments of a receiver unit capable of processing a data transmission, determining the characteristics of the transmission channels (e.g., the post-processed SNR), and reporting CSI back to the transmitter unit. Other designs based on the techniques presented herein and other receiver processing techniques may also be contemplated and are within the scope of the invention.

Channel State Information (CSI)

[1257] The CSI used to select the proper data rate and the coding and modulation scheme for each independent data stream may comprise any type of information that is indicative of the characteristics of the communication link. The CSI may be categorized as either "full CSI" or "partial CSI". Various types of information may be provided as full or partial CSI, and some examples are described below.

[1258] In one embodiment, the partial CSI comprises SNR, which may be derived as the ratio of the signal power over the noise-and-interference power. The SNR is typically estimated and provided for each transmission channel used for data transmission (e.g., each transmit data stream), although an aggregate SNR may also be provided for a number of transmission channels. The SNR estimate may be quantized to a value having a particular number of bits. In one embodiment, the SNR estimate is mapped to an SNR index, e.g., using a look-up table.

[1259] In another embodiment, the partial CSI comprises signal power and noise-and-interference power. These two components may be separately derived and provided for each transmission channel or each set of transmission channels used for data transmission.

[1260] In yet another embodiment, the partial CSI comprises signal power, noise power, and interference power. These three components may be derived and provided for each transmission channel or a set of transmission channels used for data transmission.

[1261] In yet another embodiment, the partial CSI comprises signal-to-noise ratio and a list of interference powers for observable interference terms. This information may be derived and provided for each transmission channel or each set of transmission channels used for data transmission.

[1262] In yet another embodiment, the partial CSI comprises signal components in a matrix form (e.g., $N_R \times N_T$ complex entries for all transmit-receive antenna pairs) and the noise-and-interference components in matrix form (e.g., $N_R \times N_T$ complex entries). The transmitter unit may then properly combine the signal components and the noise-and-interference components for the appropriate transmit-receive antenna pairs to derive the quality of each transmission channel used for data transmission (e.g., the post-processed SNR for each transmitted data stream, as received at the receiver unit).

[1263] In yet another embodiment, the partial CSI comprises a data rate indicator for each transmit data stream. The quality of the transmission channels to be used for data transmission may be determined initially (e.g., based on the SNR estimated for the transmission channel) and a data rate corresponding to the determined channel quality may then be identified (e.g., based on a look-up table) for each transmission channel or each group of transmission channels. The identified data rate is indicative of the maximum data rate that may be transmitted on the transmission channel for the required level of performance. The data rate may be mapped to and represented by a data rate indicator (DRI), which may be efficiently coded. For example, if (up to) seven possible data rates are supported by the transmitter unit for each transmit antenna, then a 3-bit value may be used to represent the DRI where, e.g., a zero may indicate a data rate of zero (i.e., don't use the transmit antenna) and 1 through 7 may be used to indicate seven different data rates. In a typical implementation, the channel quality measurements (e.g., the SNR estimates) are mapped directly to the DRI based on, e.g., a look-up table.

[1264] In yet another embodiment, the partial CSI comprises a rate to be used at the transmitter unit for each data stream. In this embodiment, the rate may identify the particular coding and modulation scheme to be used for the data stream such that the desired level of performance is achieved.

[1265] In yet another embodiment, the partial CSI comprises a differential indicator for a particular measure of quality for a transmission channel. Initially, the SNR or DRI or some other quality measurement for the transmission channel is determined and reported as a reference measurement value. Thereafter, monitoring of the quality of the

link continues, and the difference between the last reported measurement and the current measurement is determined. The difference may then be quantized to one or more bits, and the quantized difference is mapped to and represented by the differential indicator, which is then reported. The differential indicator may indicate an increase or decrease to the last reported measurement by a particular step size (or to maintain the last reported measurement). For example, the differential indicator may indicate that (1) the observed SNR for a particular transmission channel has increased or decreased by a particular step size, or (2) the data rate should be adjusted by a particular amount, or some other change. The reference measurement may be transmitted periodically to ensure that errors in the differential indicators and/or erroneous reception of these indicators do not accumulate.

[1266] Full CSI includes sufficient characterization (e.g., the complex gain) across the entire system bandwidth (i.e., each frequency subchannel) for the propagation path between each transmit-receive antenna pair in the $N_R \times N_T$ channel response matrix $\underline{\mathbf{H}}(k)$.

[1267] In one embodiment, the full CSI comprises eigenmodes plus any other information that is indicative of, or equivalent to, SNR. For example, the SNR-related information may be a data rate indication per eigenmode, an indication of the coding and modulation scheme to be used per eigenmode, the signal and interference power per eigenmode, the signal to interference ratio per eigenmode, and so on. The information described above for the partial CSI may also be provided as the SNR related information.

[1268] In another embodiment, the full CSI comprises a matrix $\underline{\mathbf{A}} = \underline{\mathbf{H}}^H \underline{\mathbf{H}}$. This matrix $\underline{\mathbf{A}}$ is sufficient to determine the eigenmodes and eigenvalues of the channel, and may be a more efficient representation of the channel (e.g., fewer bits may be required to transmit the full CSI for this representation).

[1269] Differential update techniques may also be used for all of the full CSI data types. For example, differential updates to the full CSI characterization may be sent periodically, when the channel changes by some amount, and so on.

[1270] Other forms of full or partial CSI may also be used and are within the scope of the invention. In general, the full or partial CSI includes sufficient information in whatever form that may be used to adjust the processing at the transmitter unit such that the desired level of performance is achieved for the transmitted data streams.

Deriving and Reporting CSI

[1271] The CSI may be derived based on the signals transmitted by the transmitter unit and received at the receiver unit. In an embodiment, the CSI is derived based on a pilot included in the transmitted signals. Alternatively or additionally, the CSI may be derived based on the data included in the transmitted signals.

[1272] In yet another embodiment, the CSI comprises one or more signals transmitted on the reverse link from the receiver unit to the transmitter unit. In some systems, a degree of correlation may exist between the downlink and uplink (e.g. for time division duplexed (TDD) systems, where the uplink and downlink share the same system bandwidth in a time division multiplexed manner). In these systems, the quality of the downlink may be estimated (to a requisite degree of accuracy) based on the quality of the uplink, which may be estimated based on signals (e.g., pilot signals) transmitted from the receiver unit. The pilot signals transmitted on the uplink would then represent a means by which the transmitter unit could estimate the CSI as observed at the receiver unit. In TDD systems, the transmitter unit can derive the channel response matrix $\underline{\mathbf{H}}(k)$ (e.g., based on the pilot transmitted on the uplink), account for differences between the transmit and receive array manifolds, and receive an estimate of the noise variance at the receiver unit. The array manifold deltas may be resolved by a periodic calibration procedure that may involve feedback between the receiver unit and transmitter unit.

[1273] The signal quality may be estimated at the receiver unit based on various techniques. Some of these techniques are described in the following patents, which are assigned to the assignee of the present application and incorporated herein by reference:

- U.S Patent No. 5,799,005, entitled "System and Method for Determining Received Pilot Power and Path Loss in a CDMA Communication System," issued August 25, 1998;
- U.S. Patent No. 5,903,554, entitled "Method and Apparatus for Measuring Link Quality in a Spread Spectrum Communication System," issued May 11, 1999;
- U.S. Patent Nos. 5,056,109, and 5,265,119, both entitled "Method and Apparatus for Controlling Transmission Power in a CDMA Cellular Mobile Telephone System," respectively issued October 8, 1991 and November 23, 1993; and

* U.S Patent No. 6,097,972, entitled "Method and Apparatus for Processing Power Control Signals in CDMA Mobile Telephone System," issued August 1, 2000.

[1274] The CSI may be reported back to the transmitter unit using various CSI transmission schemes. For example, the CSI may be sent in full, differentially, or a combination thereof. In one embodiment, full or partial CSI is reported periodically, and differential updates are sent based on the prior transmitted CSI. As an example for full CSI, the updates may be corrections (based on an error signal) to the reported eigenmodes. The eigenvalues typically do not change as rapidly as the eigenmodes, so these may be updated at a lower rate. In another embodiment, the CSI is sent only when there is a change (e.g., if the change exceeds a particular threshold), which may lower the effective rate of the feedback channel. As an example for partial CSI, the SNRs may be sent back (e.g., differentially) only when they change. For an OFDM system, correlation in the frequency domain may be exploited to permit reduction in the amount of CSI to be fed back. As an example for an OFDM system using partial CSI, if the SNR corresponding to a particular spatial subchannel for N_M frequency subchannels is similar, the SNR and the first and last frequency subchannels for which this condition is true may be reported. Other compression and feedback channel error recovery techniques to reduce the amount of data to be fed back for CSI may also be used and are within the scope of the invention.

[1275] Various types of information for CSI and various CSI reporting mechanisms are also described in U.S Patent Application Serial No. 08/963,386, entitled "Method and Apparatus for High Rate Packet Data Transmission," filed November 3, 1997, assigned to the assignee of the present application, and in "TIE/EIA/IS-856 cdma2000 High Rate Packet Data Air Interface Specification", both of which are incorporated herein by reference.

[1276] For clarity, various aspects and embodiments of the resource allocation have been described specifically for the downlink and uplink. Various techniques described herein may also be used to allocate resources in "ad hoc" or peer-to-peer networks, and this is within the scope of the invention.

[1277] The MIMO-OFDM system described herein may also be designed to implement any number of standards and designs for CDMA, TDMA, FDMA, and other multiple access techniques. The CDMA standards include the IS-95, cdma2000, and W-CDMA standards, and the TDMA standards include the Global System for Mobile

Communications (GSM) standard. These standards are known in the art and incorporated herein by reference.

[1278] The elements of the base station and terminals may be implemented with one or more digital signal processors (DSP), application specific integrated circuits (ASIC), processors, microprocessors, controllers, microcontrollers, field programmable gate arrays (FPGA), programmable logic devices, other electronic units, or any combination thereof. Some of the functions and processing described herein may also be implemented with software executed on a processor.

[1279] Certain aspects of the invention may be implemented with a combination of software and hardware. For example, the processing to schedule terminals for downlink and/or uplink data transmission may be performed based on program codes executed on a processor (scheduler 834 in FIG. 8).

[1280] Headings are included herein for reference and to aid in locating certain sections. These heading are not intended to limit the scope of the concepts described therein under, and these concepts may have applicability in other sections throughout the entire specification.

[1281] The previous description of the disclosed embodiments is provided to enable any person skilled in the art to make or use the present invention. Various modifications to these embodiments will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other embodiments without departing from the spirit or scope of the invention. Thus, the present invention is not intended to be limited to the embodiments shown herein but is to be accorded the widest scope consistent with the principles and novel features disclosed herein.

[1282] **WHAT IS CLAIMED IS:**

CLAIMS

1. A method for scheduling data transmission for a plurality of terminals in a wireless communication system, comprising:

forming at least one set of terminals for possible data transmission for each of a plurality of frequency bands, wherein each set includes one or more terminals and corresponds to a hypothesis to be evaluated;

evaluating the performance of each hypothesis;

selecting one hypothesis for each frequency band based on the evaluated performance; and

scheduling the one or more terminals in each selected hypothesis for data transmission on the corresponding frequency band.

2. The method of claim 1, wherein each frequency band corresponds to a respective group of one or more frequency subchannels.

3. The method of claim 1, wherein the plurality of terminals are scheduled for downlink data transmission.

4. The method of claim 3, further comprising:

forming one or more sub-hypotheses for each hypothesis, wherein each sub-hypothesis corresponds to specific assignments of a plurality of transmit antennas to the one or more terminals in the hypothesis, and wherein the performance of each sub-hypothesis is evaluated and one sub-hypothesis is selected for each frequency band based on the evaluated performance.

5. The method of claim 3, further comprising:

assigning a plurality of transmit antennas to the one or more terminals in each hypothesis, and wherein the performance of each hypothesis is evaluated based in part on the antenna assignments for the hypothesis.

6. The method of claim 5, wherein the assigning for each hypothesis includes

identifying a transmit antenna and terminal pair with a best metric among all unassigned transmit antennas,

assigning the transmit antenna in the pair to the terminal in the pair, and

removing the assigned transmit antenna and terminal from consideration for the hypothesis.

7. The method of claim 5, wherein the plurality of transmit antennas are assigned to the one or more terminals in each hypothesis based on a priority of each terminal.

8. The method of claim 7, wherein the highest priority terminal in each hypothesis is assigned a transmit antenna associated with a highest throughput, and the lowest priority terminal in the hypothesis is assigned a transmit antenna associated with a lowest throughput.

9. The method of claim 3, further comprising:

forming a channel response matrix for a plurality of terminals in a particular hypothesis, and wherein the performance of the hypothesis is evaluated based on the channel response matrix.

10. The method of claim 9, wherein the evaluating includes

deriving a matrix of steering vectors to be used to generate a plurality of beams for the plurality of terminals in the particular hypothesis.

11. The method of claim 10, further comprising:

deriving a scaling matrix to be used to adjust transmit power for each terminal in the particular hypothesis.

12. The method of claim 1, wherein the plurality of terminals are scheduled for uplink data transmission.

13. The method of claim 12, further comprising:

forming one or more sub-hypotheses for each hypothesis, wherein each sub-hypothesis corresponds to a specific ordering of the one or more terminals in the hypothesis, and wherein the performance of each sub-hypothesis is evaluated and one sub-hypothesis is selected for each frequency band based on the evaluated performance.

14. The method of claim 13, wherein one terminal ordering is formed for each hypothesis based on a priority of each terminal in the hypothesis.

15. The method of claim 13, wherein each sub-hypothesis is evaluated by

processing signals hypothetically transmitted from the one or more terminals in the sub-hypothesis to provide processed signals, and

estimating signal-to-noise-and-interference ratios (SNRs) for the processed signals.

16. The method of claim 15, wherein the SNRs for the processed signals are dependent on a particular order in which the hypothetically transmitted signals are processed, and wherein the hypothetically transmitted signals are processed in a specific order determined by the terminal ordering for the sub-hypothesis being evaluated.

17. The method of claim 15, wherein one sub-hypothesis is formed for each hypothesis, and wherein the terminal ordering for the sub-hypothesis is determined based on the SNRs for the processed signals.

18. The method of claim 15, wherein one sub-hypothesis is formed for each hypothesis, and wherein transmitted signals from a lowest priority terminal in the hypothesis are processed first and transmitted signals from a highest priority terminal are processed last.

19. The method of claim 12, wherein the performance of each hypothesis is evaluated based on successive cancellation receiver processing.

20. The method of claim 19, wherein the successive cancellation receiver processing performs a plurality of iterations to recover a plurality of signals

hypothetically transmitted from the one or more terminals in each hypothesis, one iteration for each hypothetically transmitted signal to be recovered.

21. The method of claim 20, wherein each iteration includes

processing a plurality of input signals in accordance with a particular equalization scheme to provide a plurality of processed signals,

detecting the processed signal corresponding to the hypothetically transmitted signal being recovered in the iteration to provide a decoded data stream, and

selectively deriving a plurality of modified signals based on the input signals and having interference components due to the decoded data stream approximately removed, and

wherein the input signals for a first iteration are signals received from the one or more terminals in the hypothesis being evaluated and the input signals for each subsequent iteration are the modified signals from a preceding iteration.

22. The method of claim 1, wherein each hypothesis is evaluated based in part on channel state information (CSI) for each terminal in the hypothesis.

23. The method of claim 22, wherein the channel state information comprises signal-to-noise-and-interference ratios (SNRs).

24. The method of claim 23, wherein each set of one or more terminals to be evaluated for a particular frequency band is associated with a respective matrix of SNRs achieved by the one or more terminals in the set for that frequency band.

25. The method of claim 22, wherein the channel state information comprises a channel gain for each transmit-receive antenna pair to be used for data transmission.

26. The method of claim 1, further comprising:

determining a data rate for each data stream to be transmitted for each scheduled terminal, and wherein a plurality of data steams are transmitted at the determined data rates.

27. The method of claim 26, further comprising:

determining a coding and modulation scheme to be used for each data stream to be transmitted, and wherein the plurality of data streams are processed based on the determined coding and modulation schemes prior to transmission.

28. The method of claim 1, wherein the plurality of terminals are scheduled for data transmission over a plurality of spatial subchannels.

29. The method of claim 28, wherein each selected hypothesis includes a plurality of SIMO terminals, wherein each SIMO terminal is assigned one spatial subchannel.

30. The method of claim 28, wherein each selected hypothesis includes a single MIMO terminal assigned all spatial subchannels.

31. The method of claim 28, wherein each selected hypothesis includes a combination of SIMO and MIMO terminals, wherein each SIMO terminal is assigned one spatial subchannel and each MIMO terminal is assigned two or more spatial subchannels.

32. The method of claim 1, wherein at least one set includes a plurality of MISO terminals each having a single antenna to receive a downlink data transmission.

33. The method of claim 1, wherein each set of multiple terminals includes terminals having similar link margins.

34. The method of claim 1, wherein the evaluating for each hypothesis includes

computing a performance metric for the hypothesis.

35. The method of claim 34, wherein the performance metric is a function of an overall throughput achievable by the one or more terminals in the hypothesis for a particular frequency band.

36. The method of claim 35, wherein the throughput for each terminal in the hypothesis is determined based on a signal-to-noise-and-interference ratio (SNR) achieved by the terminal for the particular frequency band.

37. The method of claim 35, wherein the throughput for each terminal is determined based on a signal-to-noise-and-interference ratio (SNR) achieved by the terminal for each of a plurality of frequency subchannels in the particular frequency band.

38. The method of claim 34, wherein for each frequency band the hypothesis having the best performance metric is selected for scheduling.

39. The method of claim 1, further comprising:

prioritizing the plurality of terminals to be scheduled for data transmission.

40. The method of claim 39, further comprising:

selecting a group of N highest priority terminals to be considered for scheduling for each frequency band, where N is one or greater.

41. The method of claim 39, further comprising:

maintaining one or more metrics for each terminal to be considered for scheduling, and wherein the priority of each terminal is determined based on the one or more metrics maintained for the terminal.

42. The method of claim 41, wherein one metric maintained for each terminal relates to an average throughput achieved by the terminal.

43. The method of claim 39, wherein the priority of each terminal is determined based on one or more factors maintained for the terminal and associated with quality of service (QoS).

44. In a multiple-input multiple-output (MIMO) communication system utilizing orthogonal frequency division multiplexing (OFDM), a method for scheduling downlink data transmission for a plurality of terminals, comprising:

forming at least one set of terminals for possible data transmission for each of a plurality of frequency bands, wherein each set includes one or more terminals and corresponds to a hypothesis to be evaluated, and wherein each frequency band corresponds to a respective group of one or more frequency subchannels;

forming one or more sub-hypotheses for each hypothesis, wherein each sub-hypothesis corresponds to specific assignments of a plurality of transmit antennas to the one or more terminals in the hypothesis;

evaluating the performance of each sub-hypothesis;

selecting one sub-hypothesis for each frequency band based on the evaluated performance; and

scheduling the one or more terminals in each selected sub-hypothesis for downlink data transmission on the corresponding frequency band.

45. The method of claim 44, wherein the evaluating for each sub-hypothesis includes

determining an overall throughput for the one or more terminals in the sub-hypothesis based on the specific antenna assignments, and

wherein for each frequency band the sub-hypothesis with the highest throughput is selected.

46. The method of claim 44, wherein one set of terminals is formed, and wherein the one or more terminals in each set are selected based on priority.

47. In a multiple-input multiple-output (MIMO) communication system utilizing orthogonal frequency division multiplexing (OFDM), a method for scheduling downlink data transmission for a plurality of terminals, comprising:

forming at least one set of terminals for possible data transmission for each of a plurality of frequency bands, wherein each set includes a plurality of terminals and corresponds to a hypothesis to be evaluated, and wherein each frequency band corresponds to a respective group of one or more frequency subchannels;

forming a channel response matrix for the plurality of terminals in each hypothesis;

evaluating the performance of each hypothesis based on the channel response matrix;

selecting one hypothesis for each frequency band based on the evaluated performance; and

scheduling the one or more terminals in each selected hypothesis for downlink data transmission on the corresponding frequency band.

48. In a multiple-input multiple-output (MIMO) communication system utilizing orthogonal frequency division multiplexing (OFDM), a method for scheduling uplink data transmission for a plurality of terminals, comprising:

forming at least one set of terminals for possible data transmission for each of a plurality of frequency bands, wherein each set includes one or more terminals and corresponds to a hypothesis to be evaluated, and wherein each frequency band corresponds to a respective group of one or more frequency subchannels;

forming one or more sub-hypotheses for each hypothesis, wherein each sub-hypothesis corresponds to a specific ordering of the one or more terminals in the hypothesis

evaluating the performance of each sub-hypothesis;

selecting one sub-hypothesis for each frequency band based on the evaluated performance; and

scheduling the one or more terminals in each selected sub-hypothesis for uplink data transmission on the corresponding frequency band.

49. The method of claim 48, wherein signals transmitted from the one or more scheduled terminals in the selected sub-hypothesis for each frequency band are processed in a particular order determined by the ordering for the sub-hypothesis.

50. The method of claim 48, wherein the evaluating for each sub-hypothesis includes

processing each signal hypothetically transmitted from each terminal in the sub-hypothesis to provide a corresponding processed signal, and

determining a signal-to-noise-and-interference ratio (SNR) for each processed signal.

51. The method of claim 50, wherein one sub-hypothesis is formed for each hypothesis, and wherein the ordering in the sub-hypothesis is selected to achieve a best performance for the hypothesis, as determined by one or more performance metrics.

52. A memory communicatively coupled to a digital signal processing device (DSPD) capable of interpreting digital information to:

receive channel state information (CSI) indicative of channel estimates for a plurality of terminals in a wireless communication system;

form at least one set of terminals for possible data transmission for each of a plurality of frequency bands, wherein each set includes one or more terminals and corresponds to a hypothesis to be evaluated;

evaluate the performance of each hypothesis based in part on the channel state information for the one or more terminals in the hypothesis;

select one hypothesis for each frequency band based on the evaluated performance; and

schedule the one or more terminals in each selected hypothesis for data transmission on the corresponding frequency band.

53. A computer program product for scheduling data transmission for a plurality of terminals in a wireless communication system, comprising:

code for receiving channel state information (CSI) indicative of channel estimates for a plurality of terminals in the communication system;

code for forming at least one set of terminals for possible data transmission for each of a plurality of frequency bands, wherein each set includes one or more terminals and corresponds to a hypothesis to be evaluated;

code for evaluating the performance of each hypothesis based in part on the channel state information for the one or more terminals in the hypothesis;

code for selecting one hypothesis for each frequency band based on the evaluated performance;

code for scheduling the one or more terminals in each selected hypothesis for data transmission on the corresponding frequency band; and

a computer-usable medium for storing the codes.

54. A scheduler in a multiple-input multiple-output (MIMO) communication system utilizing orthogonal frequency division multiplexing (OFDM), comprising:

means for receiving channel state information (CSI) indicative of channel estimates for a plurality of terminals in the communication system;

means for forming at least one set of terminals for possible data transmission for each of a plurality of frequency bands, wherein each set includes one or more terminals and corresponds to a hypothesis to be evaluated;

means for evaluating the performance of each hypothesis based in part on the channel state information for the one or more terminals in the hypothesis;

means for selecting one hypothesis for each frequency band based on the evaluated performance; and

means for scheduling the one or more terminals in each selected hypothesis for data transmission on the corresponding frequency band.

55. The scheduler of claim 54, further comprising:

means for forming one or more sub-hypotheses for each hypothesis, wherein each sub-hypothesis corresponds to specific assignments of a plurality of transmit antennas to the one or more terminals in the hypothesis for downlink data transmission, wherein the performance of each sub-hypothesis is evaluated and one sub-hypothesis is selected for each frequency band based on the evaluated performance.

56. The scheduler of claim 54, further comprising:

means for forming one or more sub-hypotheses for each hypothesis, wherein each sub-hypothesis corresponds to a specific order for processing uplink data transmissions from the one or more terminals in the hypothesis, wherein the

performance of each sub-hypothesis is evaluated and one sub-hypothesis is selected for each frequency band based on the evaluated performance.

57. The scheduler of claim 54, further comprising:

means for prioritizing the plurality of terminals to be scheduled for data transmission.

58. A base station in a multiple-input multiple-output (MIMO) communication system utilizing orthogonal frequency division multiplexing (OFDM), comprising:

a scheduler operative to receive channel state information (CSI) indicative of channel estimates for a plurality of terminals in the communication system, select a set of one or more terminals for data transmission for each of a plurality of frequency bands, and assign the one or more terminals in each selected set with a plurality of spatial subchannels in the corresponding frequency band;

a transmit data processor operative to receive and process data to provide a plurality of data streams for transmission to one or more scheduled terminals, wherein the data is processed based on the channel state information for the one or more scheduled terminals;

at least one modulator operative to process the plurality of data streams to provide a plurality of modulated signals; and

a plurality of antennas configured to receive and transmit the plurality of modulated signals to the one or more scheduled terminals.

59. The base station of claim 58, wherein the scheduler is further operative to select a data rate for each data stream.

60. The base station of claim 58, wherein the scheduler is further operative to select a coding and modulation scheme to be used for each data stream, and wherein the transmit data processor is further operative to process the data for each data stream based on the coding and modulation scheme selected for the data stream.

61. The base station of claim 58, further comprising:

at least one demodulator operative to process a plurality of signals received via the plurality of antennas to provide a plurality of received signals, and

a receive data processor operative to process the plurality of received signals to derive channel state information for the plurality of terminals in the communication system.

62. A transmitter apparatus in a multiple-input multiple-output (MIMO) communication system utilizing orthogonal frequency division multiplexing (OFDM), comprising:

means for receiving channel state information (CSI) indicative of channel estimates for a plurality of terminals in the communication system;

means for selecting a set of one or more terminals for data transmission for each of a plurality of frequency bands;

means for assigning the one or more terminals in each selected set with a plurality of spatial subchannels in the corresponding frequency band;

means for processing data to provide a plurality of data streams for transmission to one or more scheduled terminals, wherein the data is processed based on the channel state information for the one or more scheduled terminals;

means for processing the plurality of data streams to provide a plurality of modulated signals; and

means for transmitting the plurality of modulated signals to the one or more scheduled terminals.

63. A terminal in a multiple-input multiple-output (MIMO) communication system, comprising:

a plurality of antennas, each antenna configured to receive a plurality of transmitted signals and to provide a respective received signal;

a plurality of front-end units, each front-end unit operative to process a respective received signal to provide a corresponding stream of samples, and to derive channel state information (CSI) for a plurality of sample streams;

a receive processor operative to process the plurality of sample streams from the plurality of front-end units to provide one or more decoded data streams; and

a transmit data processor operative to process the channel state information for transmission, and

wherein the terminal is one of one or more terminals included in a set scheduled for data transmission via one or more of a plurality of frequency bands for a particular time interval.

64. The terminal of claim 63, further comprising:

at least one demodulator operative to process the plurality of sample streams to provide one or more received symbol streams for one or more spatial subchannels of one or more frequency subchannels assigned to the terminal for downlink data transmission.

65. A multiple-input multiple-output (MIMO) communication system utilizing orthogonal frequency division multiplexing (OFDM), comprising:

a scheduler operative to receive channel state information (CSI) indicative of channel estimates for a plurality of terminals in the communication system, select a set of one or more terminals for data transmission on each of a plurality of frequency bands, and assign the one or more terminals in each selected set with a plurality of spatial subchannels in the corresponding frequency band;

a base station operative to process transmissions for one or more terminals scheduled for data transmission on the plurality of spatial subchannels of the plurality of frequency bands; and

a plurality of terminals, each terminal operative to communicate with the base station via one or more spatial subchannels of one or more frequency bands assigned to the terminal when scheduled for data transmission by the scheduler.

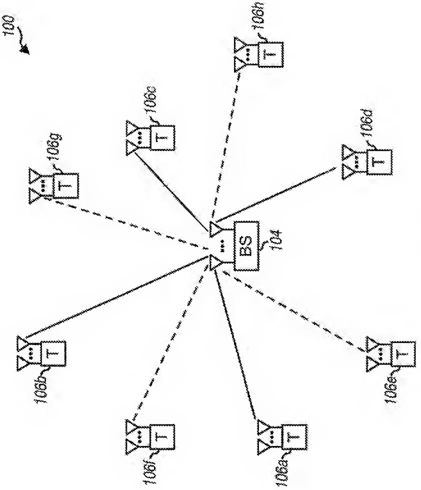


FIG. 1

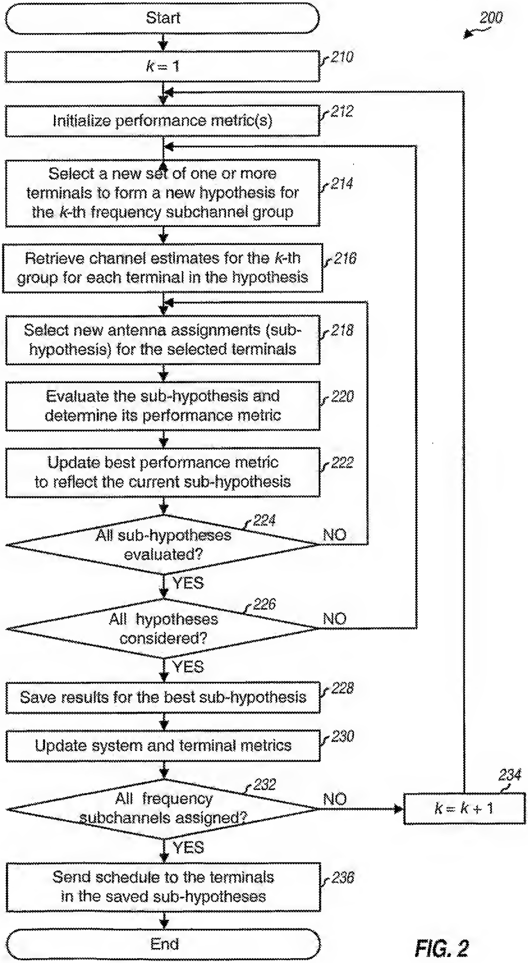

200
Start
k = 1
210
Initialize performance metric(s)
212
Select a new set of one or more terminals to form a new hypothesis for the k-th frequency subchannel group
214
Retrieve channel estimates for the k-th group for each terminal in the hypothesis
216
Select new antenna assignments (sub-hypothesis) for the selected terminals
218
Evaluate the sub-hypothesis and determine its performance metric
220
Update best performance metric to reflect the current sub-hypothesis
222
All sub-hypotheses evaluated?
224
NO
YES
All hypotheses considered?
226
NO
YES
Save results for the best sub-hypothesis
228
Update system and terminal metrics
230
All frequency subchannels assigned?
232
NO
k = k + 1
234
YES
Send schedule to the terminals in the saved sub-hypotheses
236
End

FIG. 2

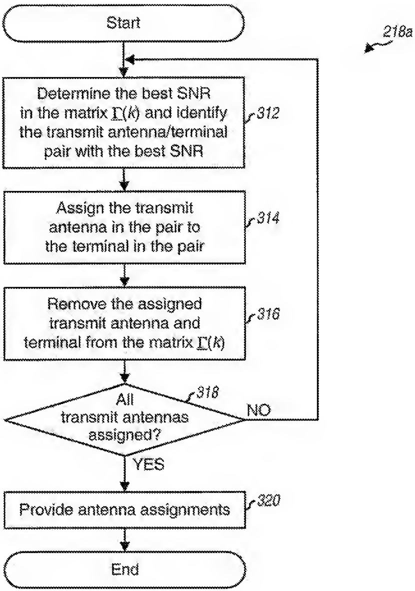

218a
Start
Determine the best SNR in the matrix Γ(k) and identify the transmit antenna/terminal pair with the best SNR
312
Assign the transmit antenna in the pair to the terminal in the pair
314
Remove the assigned transmit antenna and terminal from the matrix Γ(k)
316
All transmit antennas assigned?
318
NO
YES
Provide antenna assignments
320
End

FIG. 3

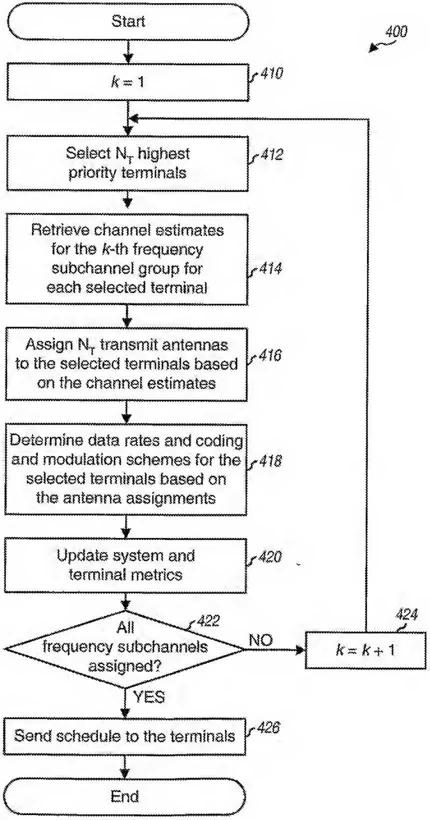

Start
400
$k = 1$
410
Select N_T highest priority terminals
412
Retrieve channel estimates for the k-th frequency subchannel group for each selected terminal
414
Assign N_T transmit antennas to the selected terminals based on the channel estimates
416
Determine data rates and coding and modulation schemes for the selected terminals based on the antenna assignments
418
Update system and terminal metrics
420
All frequency subchannels assigned?
422
NO
$k = k + 1$
424
YES
Send schedule to the terminals
426
End

FIG. 4

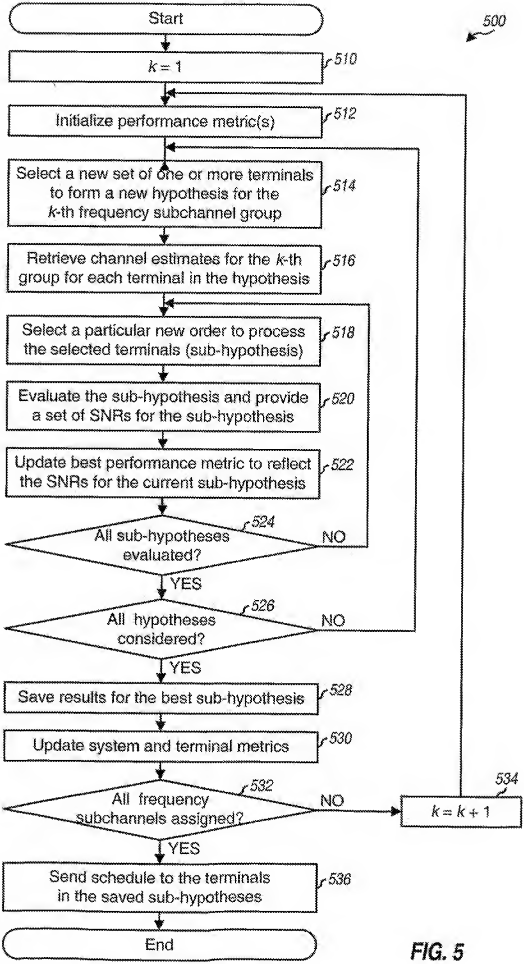

500
Start
k = 1
510
Initialize performance metric(s)
512
Select a new set of one or more terminals to form a new hypothesis for the k-th frequency subchannel group
514
Retrieve channel estimates for the k-th group for each terminal in the hypothesis
516
Select a particular new order to process the selected terminals (sub-hypothesis)
518
Evaluate the sub-hypothesis and provide a set of SNRs for the sub-hypothesis
520
Update best performance metric to reflect the SNRs for the current sub-hypothesis
522
All sub-hypotheses evaluated?
524
NO
YES
All hypotheses considered?
526
NO
YES
Save results for the best sub-hypothesis
528
Update system and terminal metrics
530
All frequency subchannels assigned?
532
NO
k = k + 1
534
YES
Send schedule to the terminals in the saved sub-hypotheses
536
End

FIG. 5

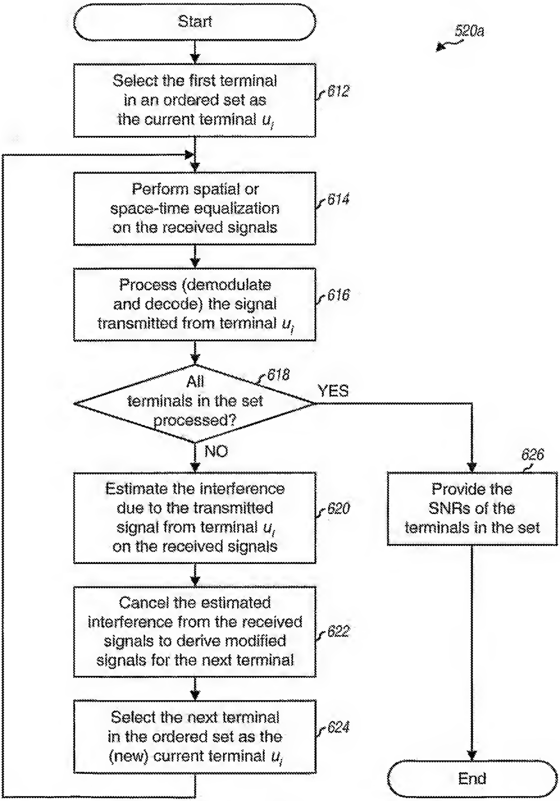

Start
520a
Select the first terminal in an ordered set as the current terminal u_i
612
Perform spatial or space-time equalization on the received signals
614
Process (demodulate and decode) the signal transmitted from terminal u_i
616
All terminals in the set processed?
618
YES
NO
Estimate the interference due to the transmitted signal from terminal u_i on the received signals
620
Cancel the estimated interference from the received signals to derive modified signals for the next terminal
622
Select the next terminal in the ordered set as the (new) current terminal u_i
624
Provide the SNRs of the terminals in the set
626
End

FIG. 6A

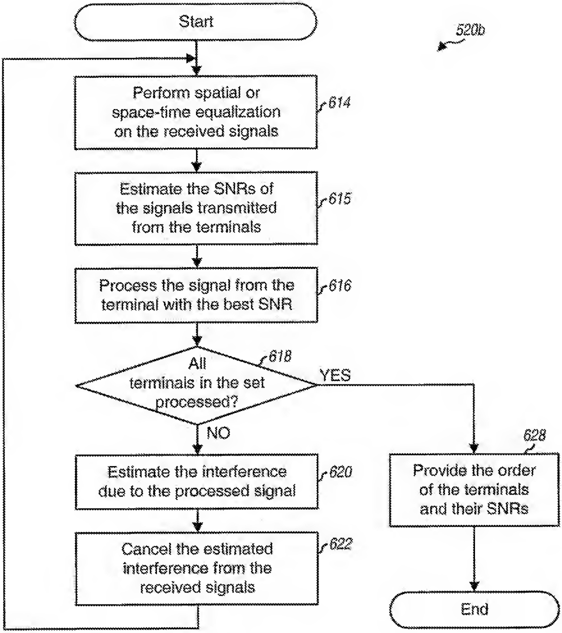

520b
Start
Perform spatial or space-time equalization on the received signals
614
Estimate the SNRs of the signals transmitted from the terminals
615
Process the signal from the terminal with the best SNR
616
All terminals in the set processed?
618
YES
NO
Estimate the interference due to the processed signal
620
Cancel the estimated interference from the received signals
622
Provide the order of the terminals and their SNRs
628
End

FIG. 6B

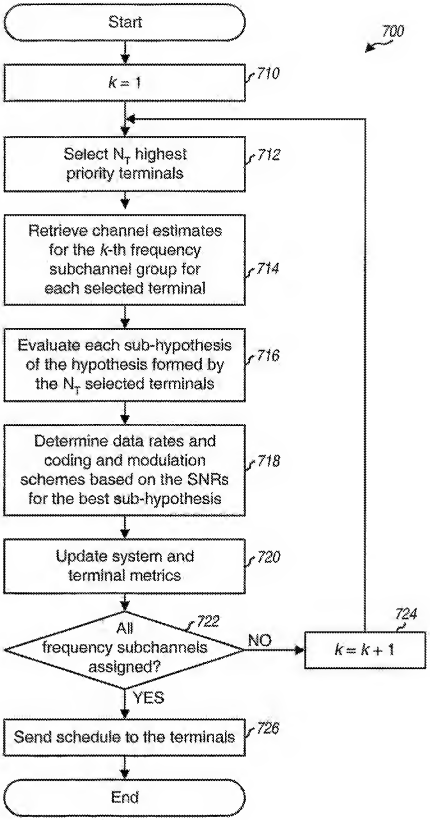

Start
700
k = 1
710
Select N_T highest priority terminals
712
Retrieve channel estimates for the k-th frequency subchannel group for each selected terminal
714
Evaluate each sub-hypothesis of the hypothesis formed by the N_T selected terminals
716
Determine data rates and coding and modulation schemes based on the SNRs for the best sub-hypothesis
718
Update system and terminal metrics
720
All frequency subchannels assigned?
722
NO
k = k + 1
724
YES
Send schedule to the terminals
726
End

FIG. 7

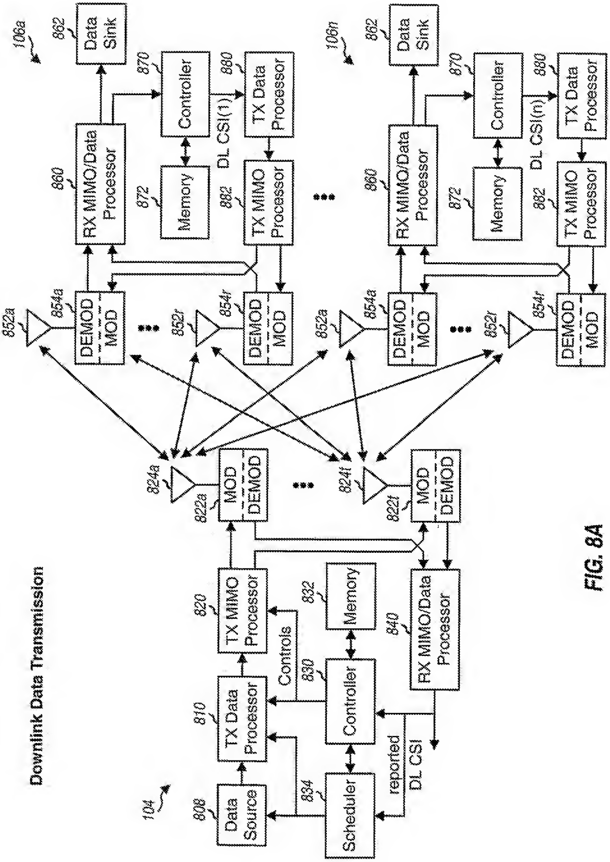

Downlink Data Transmission
104
808
Data Source
810
TX Data Processor
820
TX MIMO Processor
834
Scheduler
830
Controller
Controls
832
Memory
840
RX MIMO/Data Processor
reported DL CSI
822a
MOD
DEMOD
824a
822t
MOD
DEMOD
824t
106a
852a
854a
DEMOD
MOD
852r
854r
DEMOD
MOD
860
RX MIMO/Data Processor
862
Data Sink
870
Controller
872
Memory
880
TX Data Processor
DL CSI(1)
882
TX MIMO Processor
106n
852a
854a
DEMOD
MOD
852r
854r
DEMOD
MOD
860
RX MIMO/Data Processor
862
Data Sink
870
Controller
872
Memory
880
TX Data Processor
DL CSI(n)
882
TX MIMO Processor

FIG. 8A

Uplink Data Transmission

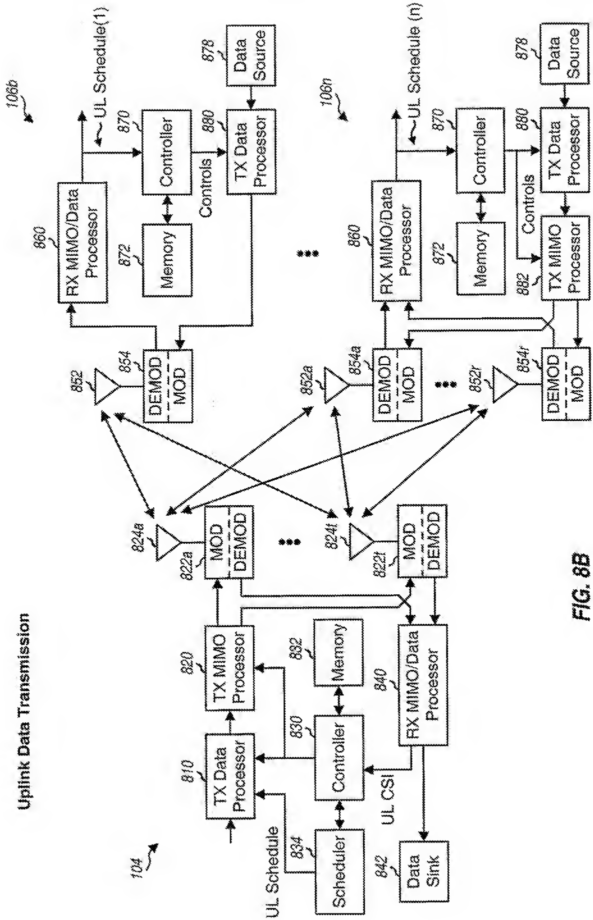


FIG. 8B

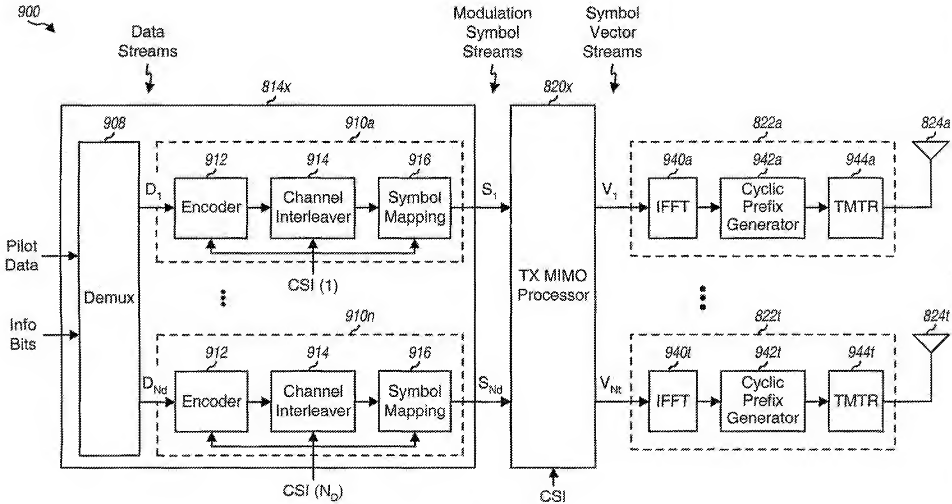

900
Data Streams
Modulation Symbol Streams
Symbol Vector Streams
814x
820x
908
910a
910n
912
914
916
Pilot Data
Info Bits
Demux
D_1
D_{Nd}
Encoder
Channel Interleaver
Symbol Mapping
CSI (1)
CSI (N_D)
S_1
S_{Nd}
TX MIMO Processor
CSI
V_1
V_{Nt}
822a
822t
940a
942a
944a
940t
942t
944t
IFFT
Cyclic Prefix Generator
TMTR
824a
824t

FIG. 9

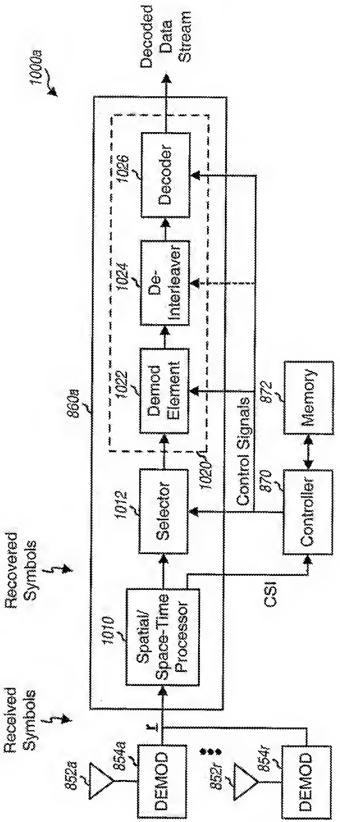

1000a
Decoded Data Stream
860a
1020
1010
Spatial/ Space-Time Processor
1012
Selector
1022
Demod Element
1024
De-Interleaver
1026
Decoder
Received Symbols
Recovered Symbols
r
852a
854a
DEMOD
852r
854r
DEMOD
CSI
870
Controller
872
Memory
Control Signals

FIG. 10A

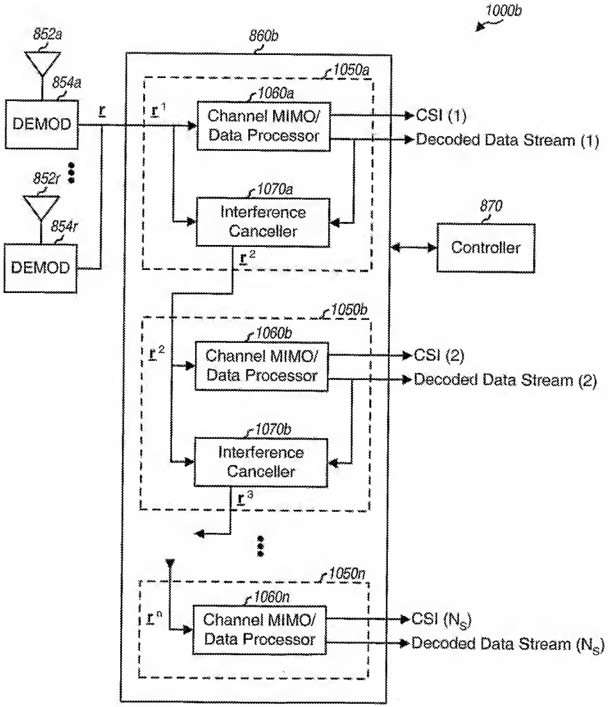

1000b
852a
854a
DEMOD
r
852r
854r
DEMOD
860b
1050a
r 1
1060a
Channel MIMO/ Data Processor
CSI (1)
Decoded Data Stream (1)
1070a
Interference Canceller
r 2
870
Controller
1050b
r 2
1060b
Channel MIMO/ Data Processor
CSI (2)
Decoded Data Stream (2)
1070b
Interference Canceller
r 3
1050n
r n
1060n
Channel MIMO/ Data Processor
CSI (N_S)
Decoded Data Stream (N_S)

FIG. 10B

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 December 2003 (31.12.2003)

PCT

(10) International Publication Number
WO 2004/002047 A1

(51) International Patent Classification[7]: H04L 1/06, 25/03, 25/02

(21) International Application Number: PCT/US2003/019464

(22) International Filing Date: 20 June 2003 (20.06.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/179,442 24 June 2002 (24.06.2002) US

(71) Applicant: QUALCOMM, INCORPORATED [US/US]; 5775 Morehouse Drive, San Diego, CA 92121 (US).

(72) Inventors: **KETCHUM, John W.**; 37 Candleberry Lane, Harvard, MA 01451 (US). **WALTON, Jay R.**; 7 Ledgewood Drive, Westford, MA 01886 (US).

(74) Agents: **WADSWORTH, Philip R.** et al.; 5775 Morehouse Drive, San Diego, CA 92121 (US).

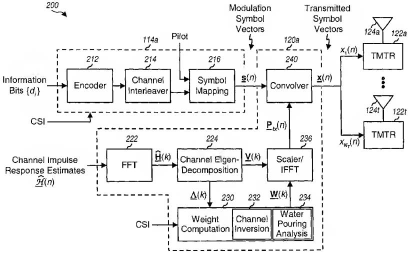
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) **Title:** SIGNAL PROCESSING WITH CHANNEL EIGENMODE DECOMPOSITION AND CHANNEL INVERSION FOR MIMO SYSTEMS



(57) **Abstract:** Techniques for processing a data transmission at a transmitter and receiver, which use channel eigen-decomposition, channel inversion, and (optionally) "water-pouring". At the transmitter, (1) channel eigen-decomposition is performed to determine eigenmodes of a MIMO channel and to derive a first set of steering vectors, (2) channel inversion is performed to derive weights (*e.g.*, one set for each eigenmode) used to minimize ISI distortion, and (3) water-pouring may be performed to derive scaling values indicative of the transmit powers allocated to the eigenmodes. The first set of steering vectors, weights, and scaling values are used to derive a pulse-shaping matrix, which is used to precondition modulation symbols prior to transmission. At the receiver, channel eigen-decomposition is performed to derive a second set of steering vectors, which are used to derive a pulse-shaping matrix used to condition received symbols such that orthogonal symbol streams are recovered.

WO 2004/002047 A1

SIGNAL PROCESSING WITH CHANNEL EIGENMODE DECOMPOSITION AND CHANNEL INVERSION FOR MIMO SYSTEMS

BACKGROUND

Field

[1001] The present invention relates generally to data communication, and more specifically to techniques for performing signal processing with channel eigenmode decomposition and channel inversion for multiple-input multiple-output (MIMO) communication systems.

Background

[1002] A multiple-input multiple-output (MIMO) communication system employs multiple (N_T) transmit antennas and multiple (N_R) receive antennas for data transmission. A MIMO channel formed by the N_T transmit and N_R receive antennas may be decomposed into N_S independent channels, with $N_S \leq \min\{N_T, N_R\}$. Each of the N_S independent channels is also referred to as a spatial subchannel of the MIMO channel and corresponds to a dimension. The MIMO system can provide improved performance (e.g., increased transmission capacity) if the additional dimensionalities created by the multiple transmit and receive antennas are utilized.

[1003] The spatial subchannels of a wideband MIMO system may encounter different channel conditions due to various factors such as fading and multipath. Each spatial subchannel may thus experience frequency selective fading, which is characterized by different channel gains at different frequencies (i.e., different frequency bins or subbands) of the overall system bandwidth. With frequency selective fading, each spatial subchannel may achieve different signal-to-noise-and-interference ratios (SNRs) for different frequency bins. Consequently, the number of information bits per modulation symbol (or data rate) that may be transmitted at different frequency bins of each spatial subchannel for a particular level of performance (e.g., 1% packet error rate) may be different from bin to bin. Moreover, because the channel conditions

typically vary with time, the supported data rates for the bins of the spatial subchannels also vary with time.

[1004] To combat frequency selective fading in a wideband channel, orthogonal frequency division multiplexing (OFDM) may be used to effectively partition the system bandwidth into a number of (N_F) subbands (which may also be referred to as frequency bins or subchannels). With OFDM, each frequency subchannel is associated with a respective subcarrier upon which data may be modulated. For a MIMO system that utilizes OFDM (i.e., a MIMO-OFDM system), each frequency subchannel of each spatial subchannel may be viewed as an independent transmission channel.

[1005] A key challenge in a coded communication system is the selection of the appropriate data rates and coding and modulation schemes to be used for a data transmission based on channel conditions. The goal of this selection process is to maximize throughput while meeting quality objectives, which may be quantified by a particular packet error rate (PER), certain latency criteria, and so on.

[1006] One straightforward technique for selecting data rates and coding and modulation schemes is to "bit load" each transmission channel in the MIMO-OFDM system according to its transmission capability, which may be quantified by the channel's short-term average SNR. However, this technique has several major drawbacks. First, coding and modulating individually for each transmission channel can significantly increase the complexity of the processing at both the transmitter and receiver. Second, coding individually for each transmission channel may greatly increase coding and decoding delay. And third, a high feedback rate would be needed to send channel state information (CSI) indicative of the channel conditions (e.g., the gain, phase, and SNR) of each transmission channel.

[1007] For a MIMO system, transmit power is another parameter that may be manipulated to maximize throughput. In general, the overall throughput of the MIMO system may be increased my allocating more transmit power to transmission channels with greater transmission capabilities. However, allocating different amounts of transmit power to different frequency bins of a given spatial subchannel tends to exaggerate the frequency selective nature of the spatial subchannel. It is well known that frequency selective fading causes inter-symbol interference (ISI), which is a phenomenon whereby each symbol in a received signal acts as distortion to subsequent symbols in the received signal. The ISI distortion degrades performance by impacting

the ability to correctly detect the received symbols. To mitigate the deleterious effects of ISI, equalization of the received symbols would need to be performed at the receiver. Thus, a major drawback in frequency-domain power allocation is the additional complexity at the receiver to combat the resultant additional ISI distortion.

[1008] There is therefore a need in the art for techniques to achieve high overall throughput in a MIMO system without having to individually code each transmission channel and which mitigate the deleterious effects of ISI.

SUMMARY

[1009] Techniques are provided herein for processing a data transmission at a transmitter and a receiver of a MIMO system such that high performance (e.g., high overall throughput) is achieved. In an aspect, a time-domain implementation is provided which uses frequency-domain channel eigen-decomposition, channel inversion, and (optionally) "water-pouring" results to derive pulse-shaping and beam-steering solutions for the transmitter and receiver.

[1010] Channel eigen-decomposition is performed at the transmitter to determine the eigenmodes (i.e., the spatial subchannels) of a MIMO channel and to obtain a first set of steering vectors, which are used to precondition modulation symbols prior to transmission over the MIMO channel. Channel eigen-decomposition may be performed based on an estimated channel response matrix, which is an estimate of the (time-domain or frequency-domain) channel response of the MIMO channel. Channel eigen-decomposition is also performed at the receiver to obtain a second set of steering vectors, which are used to condition received symbols such that orthogonal symbol streams are recovered at the receiver.

[1011] Channel inversion is performed at the transmitter to derive weights, which are used to minimize or reduce the amount of ISI distortion at the receiver. In particular, the channel inversion may be performed for each eigenmode used for data transmission. One set of weights may be derived for each eigenmode based on the estimated channel response matrix for the MIMO channel and is used to invert the frequency response of the eigenmode.

[1012] Water-pouring analysis may (optionally) be used to more optimally allocate the total available transmit power to the eigenmodes of the MIMO channel. In particular, eigenmodes with greater transmission capabilities may be allocated more

transmit power, and eigenmodes with transmission capabilities below a particular threshold may be omitted from use (e.g., by allocating these bad eigenmodes with zero transmit power). The transmit power allocated to each eigenmode then determines the data rate and possibly the coding and modulation scheme to be used for the eigenmode.

[1013] At the transmitter, data is initially processed (e.g., coded and modulated) in accordance with a particular processing scheme to provide a number of modulation symbol streams (e.g., one modulation symbol stream for each eigenmode). An estimated channel response matrix for the MIMO channel is obtained (e.g., from the receiver) and decomposed (e.g., in the frequency domain, using channel eigen-decomposition) to obtain a set of matrices of right eigen-vectors and a set of matrices of singular values. A number of sets of weights are then derived based on the matrices of singular values, with each set of weights being used to invert the frequency response of a respective eigenmode used for data transmission. Water-pouring analysis may also be performed based on the matrices of singular values to obtain a set of scaling values indicative of the transmit powers allocated to the eigenmodes. A pulse-shaping matrix for the transmitter is then derived based on the matrices of right eigen-vectors, the weights, and the scaling values (if available). The pulse-shaping matrix comprises steering vectors, which are used to precondition the streams of modulation symbols to obtain streams of preconditioned symbols to be transmitted over the MIMO channel.

[1014] At the receiver, the estimated channel response matrix is also obtained (e.g., based on pilot symbols sent from the transmitter) and decomposed to obtain a set of matrices of left eigen-vectors. A pulse-shaping matrix for the receiver is then derived based on the matrices of left eigen-vectors and used to condition a number of received symbol streams to obtain a number of recovered symbol streams. The recovered symbols are further processed (e.g., demodulated and decoded) to recover the transmitted data.

[1015] Various aspects and embodiments of the invention are described in further detail below. The invention further provides methods, digital signal processors, transmitter and receiver units, and other apparatuses and elements that implement various aspects, embodiments, and features of the invention, as described in further detail below.

BRIEF DESCRIPTION OF THE DRAWINGS

[1016] The features, nature, and advantages of the present invention will become more apparent from the detailed description set forth below when taken in conjunction with the drawings in which like reference characters identify correspondingly throughout and wherein:

[1017] FIG. 1 is a block diagram of an embodiment of a transmitter system and a receiver system in a MIMO system;

[1018] FIG. 2 is a block diagram of a transmitter unit within the transmitter system;

[1019] FIGS. 3A and 3B are diagrams that graphically illustrate the derivation of the weights used to invert the frequency response of each eigenmode of a MIMO channel;

[1020] FIG. 4 is a flow diagram of a process for allocating the total available transmit power to the eigenmodes of the MIMO channel;

[1021] FIGS. 5A and 5B are diagrams that graphically illustrate the allocation of the total transmit power to three eigenmodes in an example MIMO system;

[1022] FIG. 6 is a flow diagram of an embodiment of the signal processing at the transmitter unit;

[1023] FIG. 7 is a block diagram of a receiver unit within the receiver system; and

[1024] FIG. 8 is a flow diagram of an embodiment of the signal processing at the receiver unit.

DETAILED DESCRIPTION

[1025] The techniques described herein for processing a data transmission at a transmitter and receiver may be used for various wireless communication systems. For clarity, various aspects and embodiments of the invention are described specifically for a multiple-input multiple-output (MIMO) communication system.

[1026] A MIMO system employs multiple (N_T) transmit antennas and multiple (N_R) receive antennas for data transmission. A MIMO channel formed by the N_T transmit and N_R receive antennas may be decomposed into N_S independent channels, with $N_S \leq \min\{N_T, N_R\}$. Each of the N_S independent channels is also referred to as a spatial subchannel of the MIMO channel. The number of spatial subchannels is determined by the number of eigenmodes for the MIMO channel, which in turn is dependent on a

channel response matrix that describes the response between the N_T transmit and N_R receive antennas.

[1027] **FIG. 1** is a block diagram of an embodiment of a transmitter system 110 and a receiver system 150, which are capable of implementing various signal processing techniques described herein.

[1028] At transmitter system 110, traffic data is provided from a data source 112 to a transmit (TX) data processor 114, which formats, codes, and interleaves the traffic data based on one or more coding schemes to provide coded data. The coded traffic data may then be multiplexed with pilot data using, for example, time division multiplex (TDM) or code division multiplex (CDM), in all or a subset of the data streams to be transmitted. The pilot data is typically a known data pattern processed in a known manner, if at all. The multiplexed pilot and coded traffic data is interleaved and then modulated (i.e., symbol mapped) based on one or more modulation schemes to provide modulation symbols. In an embodiment, TX data processor 114 provides one modulation symbol stream for each spatial subchannel used for data transmission. The data rate, coding, interleaving, and modulation for each modulation symbol stream may be determined by controls provided by a controller 130.

[1029] The modulation symbols are then provided to a TX MIMO processor 120 and further processed. In a specific embodiment, the processing by TX MIMO processor 120 includes (1) determining an estimated channel frequency response matrix for the MIMO channel, (2) decomposing this matrix to determine the eigenmodes of the MIMO channel and to derive a set of "steering" vectors for the transmitter, one vector for the modulation symbol stream to be transmitted on each spatial subchannel, (3) deriving a transmit spatio-temporal pulse-shaping matrix based on the steering vectors and a weighting matrix indicative of the transmit powers assigned to the frequency bins of the eigenmodes, and (4) preconditioning the modulation symbols with the pulse-shaping matrix to provide preconditioned modulation symbols. The processing by TX MIMO processor 120 is described in further detail below. Up to N_T streams of preconditioned symbols are then provided to transmitters (TMTR) 122a through 122t.

[1030] Each transmitter 122 converts a respective preconditioned symbol stream into one or more analog signals and further conditions (e.g., amplifies, filters, and frequency upconverts) the analog signals to generate a modulated signal suitable for

transmission over the MIMO channel. The modulated signal from each transmitter 122 is then transmitted via a respective antenna 124 to the receiver system.

[1031] At receiver system 150, the transmitted modulated signals are received by N_R antennas 152a through 152r, and the received signal from each antenna 152 is provided to a respective receiver (RCVR) 154. Each receiver 154 conditions (e.g., filters, amplifies, and frequency downconverts) the received signal, digitizes the conditioned signal to provide a stream of samples, and further processes the sample stream to provide a stream of received symbols. An RX MIMO processor 160 then receives and processes the N_R received symbol streams to provide N_T streams of recovered symbols, which are estimates of the modulation symbols transmitted from the transmitter system. In an embodiment, the processing by RX MIMO processor 160 may include (1) determining the estimated channel frequency response matrix for the MIMO channel, (2) decomposing this matrix to derive a set of steering vectors for the receiver, (3) deriving a receive spatio-temporal pulse-shaping matrix based on the steering vectors, and (4) conditioning the received symbols with the pulse-shaping matrix to provide the recovered symbols. The processing by RX MIMO processor 160 is described in further detail below.

[1032] A receive (RX) data processor 162 then demodulates, deinterleaves, and decodes the recovered symbols to provide decoded data, which is an estimate of the transmitted traffic data. The processing by RX MIMO processor 160 and RX data processor 162 is complementary to that performed by TX MIMO processor 120 and TX data processor 114, respectively, at transmitter system 110.

[1033] RX MIMO processor 160 may further derive channel impulse responses for the MIMO channel, received noise power and/or signal-to-noise-and-interference ratios (SNRs) for the spatial subchannels, and so on. RX MIMO processor 160 would then provide these quantities to a controller 170. RX data processor 162 may also provide the status of each received packet or frame, one or more other performance metrics indicative of the decoded results, and possibly other information. Controller 170 then derives channel state information (CSI), which may comprise all or some of the information received from RX MIMO processor 160 and RX data processor 162. The CSI is processed by a TX data processor 178, modulated by a modulator 180, conditioned by transmitters 154a through 154r, and sent back to transmitter system 110.

[1034] At transmitter system 110, the modulated signals from receiver system 150 are received by antennas 124, conditioned by receivers 122, and demodulated by a demodulator 140 to recover the CSI transmitted by the receiver system. The CSI is then provided to controller 130 and used to generate various controls for TX data processor 114 and TX MIMO processor 120.

[1035] Controllers 130 and 170 direct the operation at the transmitter and receiver systems, respectively. Memories 132 and 172 provide storage for program codes and data used by controllers 130 and 170, respectively.

[1036] Techniques are provided herein for achieving high performance (e.g., high overall system throughput) via a time-domain implementation that uses frequency-domain channel eigen-decomposition, channel inversion, and (optionally) water-pouring results to derive time-domain pulse-shaping and beam-steering solutions for the transmitter and receiver.

[1037] Channel eigen-decomposition is performed at the transmitter to determine the eigenmodes of the MIMO channel and to derive a first set of steering vectors, which are used to precondition the modulation symbols. Channel eigen-decomposition is also performed at the receiver to derive a second set of steering vectors, which are used to condition the received symbols such that orthogonal symbol streams are recovered at the receiver. The preconditioning at the transmitter and the conditioning at the receiver orthogonalize the symbol streams transmitted over the MIMO channel.

[1038] Channel inversion is performed at the transmitter to flatten the frequency response of each eigenmode (or spatial subchannel) used for data transmission. As noted above, frequency selective fading causes intersymbol interference (ISI), which can degrade performance by impacting the ability to correctly detect the received symbols at the receiver. Conventionally, the frequency selective fading may be compensated for at the receiver by performing equalization on the received symbol streams. However, equalization increases the complexity of the receiver processing. With the inventive techniques, the channel inversion is performed at the transmitter to account for the frequency selective fading and to mitigate the need for equalization at the receiver.

[1039] Water-pouring (or water-filling) analysis is used to more optimally allocate the total available transmit power in the MIMO system to the eigenmodes such that high performance is achieved. The transmit power allocated to each eigenmode may then

determine the data rate and the coding and modulation scheme to be used for the eigenmode.

[1040] These various processing techniques are described in further detail below.

[1041] The techniques described herein provide several potential advantages. First, with time-domain eigenmode decomposition, the maximum number of eigenmodes with different SNRs is given by $\min(N_T, N_R)$. If one independent data stream is transmitted on each eigenmode and each data stream is independently processed, then the maximum number of different coding/modulation schemes is also given by $\min(N_T, N_R)$. It is also possible to make the received SNRs for the data streams essentially the same, thereby further simplifying the coding/modulation. The techniques described herein can thus greatly simplify the coding/modulation for a data transmission by avoiding the per-bin bit allocation required to approach channel capacity in MIMO-OFDM systems that utilize frequency-domain water-pouring.

[1042] Second, the channel inversion at the transmitter results in recovered symbol streams at the receiver that do not require equalization. This then reduces the complexity of the receiver processing. In contrast, other wide-band time-domain techniques typically require complicated space-time equalization to recover the symbol streams.

[1043] Third, the time-domain signaling techniques described herein can more easily integrate the channel/pilot structures of various CDMA standards, which are also based on time-domain signaling. Implementation of the channel/pilot structures may be more complicated in OFDM-based systems that perform frequency-domain signaling.

[1044] **FIG. 2** is a block diagram of an embodiment of a transmitter unit 200, which is capable of implementing various processing techniques described herein. Transmitter unit 200 is an embodiment of the transmitter portion of transmitter system 110 in FIG. 1. Transmitter unit 200 includes (1) a TX data processor 114a that receives and processes traffic and pilot data to provide N_T modulation symbol streams and (2) a TX MIMO processor 120a that preconditions the modulation symbol streams to provide N_T preconditioned symbol streams. TX data processor 114a and TX MIMO processor 120a are one embodiment of TX data processor 114 and TX MIMO processor 120, respectively, in FIG. 1.

[1045] In the specific embodiment shown in FIG. 2, TX data processor 114a includes an encoder 212, a channel interleaver 214, and a symbol mapping element 216. Encoder 212 receives and codes the traffic data (i.e., the information bits, d_i) in accordance with one or more coding schemes to provide coded bits. The coding increases the reliability of the data transmission. In an embodiment, a separate coding scheme may be used for the information bits for each eigenmode (or spatial subchannel) selected for use for data transmission. In alternative embodiments, a separate coding scheme may be used for each subset of spatial subchannels, or a common coding scheme may be used for all spatial subchannels. The coding scheme(s) to be used are determined by controls from controller 130 and may be selected based on the CSI received from the receiver system. Each selected coding scheme may include any combination of cyclic redundancy check (CRC), convolutional coding, Turbo coding, block coding, and other coding, or no coding at all.

[1046] Channel interleaver 214 interleaves the coded bits based on one or more interleaving schemes. Typically, each selected coding scheme is associated with a corresponding interleaving scheme. The interleaving provides time diversity for the coded bits, permits the data to be transmitted based on an average SNR of each spatial subchannel used for the data transmission, combats fading, and further removes correlation between coded bits used to form each modulation symbol.

[1047] Symbol mapping element 216 then receives and multiplexes pilot data with the interleaved data and further maps the multiplexed data in accordance with one or more modulation schemes to provide modulation symbols. A separate modulation scheme may be used for each spatial subchannel selected for use, or for each subset of spatial subchannels. Alternatively, a common modulation scheme may be used for all selected spatial subchannels.

[1048] The symbol mapping for each spatial subchannel may be achieved by grouping sets of bits to form data symbols (each of which may be a non-binary value) and mapping each data symbol to a point in a signal constellation corresponding to the modulation scheme selected for use for that spatial subchannel. The selected modulation scheme may be QPSK, M-PSK, M-QAM, or some other scheme. Each mapped signal point is a complex value and corresponds to a modulation symbol. Symbol mapping element 216 provides a vector of modulation symbols for each symbol

period, with the number of modulation symbols in each vector corresponding to the number of spatial subchannels selected for use for that symbol period. Symbol mapping element 216 thus provides up to N_T modulation symbol streams. These streams collectively form a sequence of vectors, with are also referred to as the modulation symbol vectors, $\underline{s}(n)$, with each such vector including up to N_S modulation symbols to be transmitted on up to N_S spatial subchannels for the n-th symbol period.

[1049] Within TX MIMO processor 120a, the response of the MIMO channel is estimated and used to precondition the modulation symbols prior to transmission to the receiver system. In a frequency division duplexed (FDD) system, the downlink and uplink are allocated different frequency bands, and the channel responses for the downlink and uplink may not be correlated to a sufficient degree. For the FDD system, the channel response may be estimated at the receiver and sent back to the transmitter. In a time division duplexed (TDD) system, the downlink and uplink share the same frequency band in a time division multiplexed manner, and a high degree of correlation may exist between the downlink and uplink channel responses. For the TDD system, the transmitter system may estimate the uplink channel response (e.g., based on the pilot transmitted by the receiver system on the uplink) and may then derive the downlink channel response by accounting for any differences such as those between the transmit and receive antenna array manifolds.

[1050] In an embodiment, the channel response estimates are provided to TX MIMO processor 120a as a sequence of $N_R \times N_T$ matrices, $\hat{\underline{\mathcal{H}}}(n)$, of time-domain samples. This sequence of matrices is collectively referred to as a channel impulse response matrix, $\hat{\underline{\underline{\mathcal{H}}}}$. The (i, j)-th element, $\underline{h}_{i,j}$, of the estimated channel impulse response matrix, $\hat{\underline{\underline{\mathcal{H}}}}$, for $i = (1, 2, \dots, N_R)$ and $j = (1, 2, \dots, N_T)$, is a sequence of samples that represents the sampled impulse response of the propagation path from the j-th transmit antenna to the i-th receive antenna.

[1051] Within TX MIMO processor 120a, a fast Fourier transformer 222 receives the estimated channel impulse response matrix, $\hat{\underline{\underline{\mathcal{H}}}}$ (e.g., from the receiver system) and derives the corresponding estimated channel frequency response matrix, $\hat{\underline{\underline{\mathbf{H}}}}$, by performing a fast Fourier transform (FFT) on the matrix $\hat{\underline{\underline{\mathcal{H}}}}$ (i.e., $\hat{\underline{\underline{\mathbf{H}}}} = \text{FFT}[\hat{\underline{\underline{\mathcal{H}}}}]$). This

may be achieved by performing an N_F-point FFT on a sequence of N_F samples for each element of $\hat{\underline{\underline{\mathcal{H}}}}$ to derive a set of N_F coefficients for the corresponding element of $\hat{\underline{\underline{\mathbf{H}}}}$, where N_F corresponds to the number of frequency bins for the FFT (i.e., the length of the FFT). The $N_R \cdot N_T$ elements of $\hat{\underline{\underline{\mathbf{H}}}}$ are thus $N_R \cdot N_T$ sets of coefficients representing the frequency responses of the propagation paths between the N_T transmit antennas and N_R receive antennas. Each element of $\hat{\underline{\underline{\mathbf{H}}}}$ is the FFT of the corresponding element of $\hat{\underline{\underline{\mathcal{H}}}}$. The estimated channel frequency response matrix, $\hat{\underline{\underline{\mathbf{H}}}}$, may also be viewed as comprising a set of N_F matrices, $\hat{\underline{\mathbf{H}}}(k)$ for $k = (0, 1, \ldots, N_F - 1)$.

Channel Eigen-Decomposition

[1052] A unit 224 then performs eigen-decomposition of the MIMO channel used for data transmission. In one embodiment for performing channel eigen-decomposition, unit 224 computes the singular value decomposition (SVD) of the estimated channel frequency response matrix, $\hat{\underline{\underline{\mathbf{H}}}}$. In an embodiment, the singular value decomposition is performed for each matrix $\hat{\underline{\mathbf{H}}}(k)$, for $k = (0, 1, \ldots, N_F - 1)$. The singular value decomposition of matrix $\hat{\underline{\mathbf{H}}}(k)$ for frequency bin k (or frequency f_k) may be expressed as:

$$\hat{\underline{\mathbf{H}}}(k) = \underline{\mathbf{U}}(k)\underline{\Lambda}(k)\underline{\mathbf{V}}^H(k) \ , \qquad \text{Eq (1)}$$

where $\underline{\mathbf{U}}(k)$ is an $N_R \times N_R$ unitary matrix (i.e., $\underline{\mathbf{U}}^H \underline{\mathbf{U}} = \underline{\mathbf{I}}$, where $\underline{\mathbf{I}}$ is the identity matrix with ones along the diagonal and zeros everywhere else);

$\underline{\Lambda}(k)$ is an $N_R \times N_T$ diagonal matrix of singular values of $\hat{\underline{\mathbf{H}}}(k)$; and

$\underline{\mathbf{V}}(k)$ is an $N_T \times N_T$ unitary matrix.

The diagonal matrix $\underline{\Lambda}(k)$ contains non-negative real values along the diagonal (i.e., $\underline{\Lambda}(k) = \text{diag}\,(\lambda_1(k), \lambda_2(k), \ldots, \lambda_{N_T}(k))$) and zeros elsewhere. The $\lambda_i(k)$, for $i = (1, 2, \ldots, N_T)$, are referred to as the singular values of the matrix $\hat{\underline{\mathbf{H}}}(k)$. The singular value decomposition is a matrix operation known in the art and described in various references. One such reference is a book by Gilbert Strang entitled "Linear

Algebra and Its Applications," Second Edition, Academic Press, 1980, which is incorporated herein by reference.

[1053] The result of the singular value decomposition is three sets of N_F matrices, $\underline{\underline{\mathbf{U}}}$, $\underline{\underline{\Lambda}}$, and $\underline{\underline{\mathbf{V}}}^H$, where $\underline{\underline{\mathbf{U}}} = [\underline{\mathbf{U}}(0) \ \ldots \ \underline{\mathbf{U}}(k) \ \ldots \ \underline{\mathbf{U}}(N_F - 1)]$, and so on. For each value of k, $\underline{\mathbf{U}}(k)$ is the $N_R \times N_R$ unitary matrix of left eigen-vectors of $\hat{\underline{\mathbf{H}}}(k)$, $\underline{\mathbf{V}}(k)$ is the $N_T \times N_T$ unitary matrix of right eigen-vectors of $\hat{\underline{\mathbf{H}}}(k)$, and $\underline{\Lambda}(k)$ is the $N_R \times N_T$ diagonal matrix of singular values of $\hat{\underline{\mathbf{H}}}(k)$.

[1054] In another embodiment for performing channel eigen-decomposition, unit 224 first obtains a square matrix $\underline{\mathbf{R}}(k)$ as $\underline{\mathbf{R}}(k) = \hat{\underline{\mathbf{H}}}^H(k)\hat{\underline{\mathbf{H}}}(k)$. The eigenvalues of the square matrix $\underline{\mathbf{R}}(k)$ would then be the squares of the singular values of the matrix $\hat{\underline{\mathbf{H}}}(k)$, and the eigen-vectors of $\underline{\mathbf{R}}(k)$ would be the right eigen-vectors of $\hat{\underline{\mathbf{H}}}(k)$, or $\underline{\mathbf{V}}(k)$. The decomposition of $\underline{\mathbf{R}}(k)$ to obtain the eigenvalues and eigen-vectors is known in the art and not described herein. Similarly, another square matrix $\underline{\mathbf{R}}'(k)$ may be obtained as $\underline{\mathbf{R}}'(k) = \hat{\underline{\mathbf{H}}}(k)\hat{\underline{\mathbf{H}}}^H(k)$. The eigenvalues of this square matrix $\underline{\mathbf{R}}'(k)$ would also be the squares of the singular values of $\hat{\underline{\mathbf{H}}}(k)$, and the eigen-vectors of $\underline{\mathbf{R}}'(k)$ would be the left eigen-vectors of $\hat{\underline{\mathbf{H}}}(k)$, or $\underline{\mathbf{U}}(k)$.

[1055] The channel eigen-decomposition is used to decompose the MIMO channel into its eigenmodes, at frequency f_k, for each value of k where $k = (0, 1, \ldots, N_F - 1)$. The rank $r(k)$ of $\hat{\underline{\mathbf{H}}}(k)$ corresponds to the number of eigenmodes for the MIMO channel at frequency f_k, which corresponds to the number of independent channels (i.e., the number of spatial subchannels) available in frequency bin k.

[1056] As described in further detail below, the columns of $\underline{\mathbf{V}}(k)$ are the steering vectors associated with frequency f_k to be used at the transmitter for the elements of the modulation symbol vectors, $\underline{\mathbf{s}}(n)$. Correspondingly, the columns of $\underline{\mathbf{U}}(k)$ are the steering vectors associated with frequency f_k to be used at the receiver for the elements of the received symbol vectors, $\underline{\mathbf{r}}(n)$. The matrices $\underline{\mathbf{U}}(k)$ and $\underline{\mathbf{V}}(k)$, for $k = (0, 1, \ldots, N_F - 1)$, are used to orthogonalize the symbol streams transmitted on the

eigenmodes at each frequency f_k. When these matrices are used to precondition the modulation symbol streams at the transmitter and to condition the received symbol streams at the receiver, either in the frequency domain or the time domain, the result is the overall orthogonalization of the symbol streams. This then allows for separate coding/modulation per eigenmode (as opposed to per bin), which can greatly simplify the processing at the transmitter and receiver.

[1057] The elements along the diagonal of $\underline{\Lambda}(k)$ are $\lambda_{ii}(k)$, for $i = \{1,\ 2,\ \ldots,\ r(k)\}$, where $r(k)$ is the rank of $\hat{\underline{\mathbf{H}}}(k)$. The columns of $\underline{\mathbf{U}}(k)$ and $\underline{\mathbf{V}}(k)$, $\underline{\mathbf{u}}_i(k)$ and $\underline{\mathbf{v}}_i(k)$, respectively, are solutions to the eigen equation, which may be expressed as:

$$\hat{\underline{\mathbf{H}}}(k)\underline{\mathbf{v}}_i(k) = \lambda_{ii}\underline{\mathbf{u}}_i(k) \quad . \qquad \text{Eq (2)}$$

[1058] The three sets of matrices, $\underline{\mathbf{U}}(k)$, $\underline{\Lambda}(k)$, and $\underline{\mathbf{V}}(k)$, for $k = (0,\ 1,\ \ldots\ ,N_F - 1)$, may be provided in two forms - a "sorted" form and a "random-ordered" form. In the sorted form, the diagonal elements of each matrix $\underline{\Lambda}(k)$ are sorted in decreasing order so that $\lambda_{11}(k) \geq \lambda_{22}(k) \geq \ldots \geq \lambda_{rr}(k)$, and their eigen-vectors are arranged in corresponding order in $\underline{\mathbf{U}}(k)$ and $\underline{\mathbf{V}}(k)$. The sorted form is indicated by the subscript *s*, i.e., $\underline{\mathbf{U}}_s(k)$, $\underline{\Lambda}_s(k)$, and $\underline{\mathbf{V}}_s(k)$, for $k = (0,\ 1,\ \ldots\ ,N_F - 1)$.

[1059] In the random-ordered form, the ordering of the singular values and eigen-vectors may be random and further independent of frequency. The random form is indicated by the subscript *r*. The particular form selected for use, sorted or random-ordered, influences the selection of the eigenmodes for use for data transmission and the coding and modulation scheme to be used for each selected eigenmode.

[1060] A weight computation unit 230 receives the set of diagonal matrices, $\underline{\underline{\Lambda}}$, which contains a set of singular values (i.e., $\lambda_{11}(k),\ \lambda_{22}(k),\ \ldots,\lambda_{rr}(k)$) for each frequency bin. Weight computation unit 230 then derives a set of weighting matrices, $\underline{\underline{\mathbf{W}}}$, where $\underline{\underline{\mathbf{W}}} = [\underline{\mathbf{W}}(0)\ \ldots\ \underline{\mathbf{W}}(k)\ \ldots\ \underline{\mathbf{W}}(N_F - 1)]$. The weighting matrices are used to scale the modulation symbol vectors, $\underline{\mathbf{s}}(n)$, in either the time or frequency domain, as described below.

[1061] Weight computation unit 230 includes a channel inversion unit 232 and (optionally) a water-pouring analysis unit 234. Channel inversion unit 232 derives a set of weights, $\underline{\mathbf{w}}_{ii}$, for each eigenmode, which is used to combat the frequency selective fading on the eigenmode. Water-pouring analysis unit 234 derives a set of scaling values, $\underline{\mathbf{b}}$, for the eigenmodes of the MIMO channel. These scaling values are indicative of the transmit powers allocated to the eigenmodes. Channel inversion and water-pouring are described in further detail below.

Channel Inversion

[1062] **FIG. 3A** is a diagram that graphically illustrates the derivation of the set of weights, $\underline{\mathbf{w}}_{ii}$, used to invert the frequency response of each eigenmode. The set of diagonal matrices, $\underline{\Lambda}(k)$ for $k = (0, 1, \ldots, N_F - 1)$, is shown arranged in order along an axis 310 that represents the frequency dimension. The singular values, $\lambda_{ii}(k)$ for $i = (1, 2, \ldots, N_S)$, of each matrix $\underline{\Lambda}(k)$ are located along the diagonal of the matrix. Axis 312 may thus be viewed as representing the spatial dimension. Each eigenmode of the MIMO channel is associated with a set of elements, $\{\lambda_{ii}(k)\}$ for $k = (0, 1, \ldots, N_F - 1)$, that is indicative of the frequency response of that eigenmode. The set of elements $\{\lambda_{ii}(k)\}$ for each eigenmode is shown by the shaded boxes along a dashed line 314. For each eigenmode that experiences frequency selective fading, the elements $\{\lambda_{ii}(k)\}$ for the eigenmode may be different for different values of k.

[1063] Since frequency selective fading causes ISI, the deleterious effects of ISI may be mitigated by "inverting" each eigenmode such that it appears flat in frequency at the receiver. The channel inversion may be achieved by deriving a set of weights, $\{w_{ii}(k)\}$ for $k = (0, 1, \ldots, N_F - 1)$, for each eigenmode such that the product of the weights and the corresponding eigenvalues (i.e., the squares of the diagonal elements) are approximately constant for all values of k, which may be expressed as $w_{ii}(k) \cdot \lambda_{ii}^2(k) = a_i$, for $k = (0, 1, \ldots, N_F - 1)$.

[1064] For eigenmode i, the set of weights for the N_F frequency bins, $\underline{\mathbf{w}}_{ii} = [w_{ii}(0) \ \ldots \ w_{ii}(k) \ \ldots \ w_{ii}(N_F - 1)]^T$, used to invert the channel may be derived as:

$$w_{ii}(k) = \frac{a_i}{\lambda_{ii}^2(k)} \quad , \qquad \text{for } k = (0,\ 1,\ \ldots\ , N_F - 1)\ , \qquad \text{Eq (3)}$$

where a_i is a normalization factor that may be expressed as:

$$a_i = \frac{\sum_{k=0}^{N_F-1} \lambda_{ii}^2(k)}{\sum_{k=0}^{N_F-1} \frac{1}{\lambda_{ii}^2(k)}} \quad . \qquad \text{Eq (4)}$$

As shown in equation (4), a normalization factor a_i is determined for each eigenmode based on the set of eigenvalues (i.e., the squared singular values), $\{\lambda_{ii}^2(k)\}$ for $k = (0,\ 1,\ \ldots\ , N_F - 1)$, associated with that eigenmode. The normalization factor a_i is defined such that $\sum_{k=0}^{N_F-1} w_{ii}(k) = \sum_{k=0}^{N_F-1} \lambda_{ii}^2(k)$.

[1065] **FIG. 3B** is a diagram that graphically illustrates the relationship between the set of weights for a given eigenmode and the set of eigenvalues for that eigenmode. For eigenmode *i*, the weight $w_{ii}(k)$ for each frequency bin is inversely related to the eigenvalue $\lambda_{ii}^2(k)$ for that bin, as shown in equation (3). To flatten the spatial subchannel and minimize or reduce ISI, it is undesirable to selectively eliminate transmit power on any frequency bin. The set of N_F weights for each eigenmode is used to scale the modulation symbols, $\underline{s}(n)$, in the frequency or time domain, prior to transmission on the eigenmode.

[1066] For the sorted order form, the singular values $\lambda_{ii}(k)$, for $i = (1,\ 2,\ \ldots\ , N_S)$, for each matrix $\underline{\Lambda}(k)$ are sorted such that the diagonal elements of $\underline{\Lambda}(k)$ with smaller indices are generally larger. Eigenmode 0 (which is often referred to as the principle eigenmode) would then be associated with the largest singular value in each of the N_F diagonal matrices, $\underline{\Lambda}(k)$, eigenmode 1 would then be associated with the second largest singular value in each of the N_F diagonal matrices, and so on. Thus, even though the channel inversion is performed over all N_F frequency bins for each eigenmode, the eigenmodes with lower indices are not likely to have too many bad bins (if any). Thus,

at least for eigenmodes with lower indices, excessive transmit power is not used for bad bins.

[1067] The channel inversion may be performed in various manners to invert the MIMO channel, and this is within the scope of the invention. In one embodiment, the channel inversion is performed for each eigenmode selected for use. In another embodiment, the channel inversion may be performed for some eigenmodes but not others. For example, the channel inversion may be performed for each eigenmode determined to induce excessive ISI. The channel inversion may also be dynamically performed for some or all eigenmodes selected for use, for example, when the MIMO channel is determined to be frequency selective (e.g., based on some defined criteria).

[1068] Channel inversion is described in further detail in U.S. Patent Application Serial No. 09/860,274, filed May 17, 2001, U.S. Patent Application Serial No. 09/881,610, filed June 14, 2001, and U.S. Patent Application Serial No. 09/892,379, filed June 26, 2001, all three entitled "Method and Apparatus for Processing Data for Transmission in a Multi-Channel Communication System Using Selective Channel Inversion," assigned to the assignee of the present application and incorporated herein by reference.

Water-Pouring

[1069] In an embodiment, water-pouring analysis is performed (if at all) across the spatial dimension such that more transmit power is allocated to eigenmodes with better transmission capabilities. The water-pouring power allocation is analogous to pouring a fixed amount of water into a vessel with an irregular bottom, where each eigenmode corresponds to a point on the bottom of the vessel, and the elevation of the bottom at any given point corresponds to the inverse of the SNR associated with that eigenmode. A low elevation thus corresponds to a high SNR and, conversely, a high elevation corresponds to a low SNR. The total available transmit power, P_{total}, is then "poured" into the vessel such that the lower points in the vessel (i.e., those with higher SNRs) are filled first, and the higher points (i.e., those with lower SNRs) are filled later. A constant P_{set} is indicative of the water surface level for the vessel after all of the total available transmit power has been poured. This constant may be estimated initially based on various system parameters. The power allocation is dependent on the total available transmit power and the depth of the vessel over the bottom surface. The

points with elevations above the water surface level are not filled (i.e., eigenmodes with SNRs below a particular value are not used for data transmission).

[1070] In an embodiment, the water-pouring is not performed across the frequency dimension because this tends to exaggerate the frequency selectivity of the eigenmodes created by the channel eigenmode decomposition described above. The water-pouring may be performed such that all eigenmodes are used for data transmission, or only a subset of the eigenmodes is used (with bad eigenmodes being discarded). It can be shown that water-pouring across the eigenmodes, when used in conjunction with the channel inversion with the singular values sorted in descending order, can provide near-optimum performance while mitigating the need for equalization at the receiver.

[1071] The water-pouring may be performed by water-pouring analysis unit 234 as follows. Initially, the total power in each eigenmode is determined as:

$$P_{i,\lambda} = \sum_{k=0}^{N_F-1} \lambda_{ii}^2(k) \quad . \qquad \text{Eq (5)}$$

[1072] The SNR for each eigenmode may then be determined as:

$$\mathrm{SNR}_i = \frac{P_{i,\lambda}}{\sigma^2} \quad , \qquad \text{Eq (6)}$$

where σ^2 is the received noise variance, which may also be denoted as the received noise power N_0. The received noise power corresponds to the noise power on the recovered symbols at the receiver, and is a parameter that may be provided by the receiver to the transmitter as part of the reported CSI.

[1073] The transmit power, P_i, to be allocated to each eigenmode may then be determined as:

$$P_i = \max\left[\left(P_{set} - \frac{1}{\mathrm{SNR}_i}\right), \ 0\right] \quad , \text{ and} \qquad \text{Eq (7a)}$$

$$P_{total} \geq \sum_{i=1}^{N_S} P_i \quad , \qquad \text{Eq (7b)}$$

where P_{set} is a constant that may be derived from various system parameters, and P_{total} is the total transmit power available for allocation to the eigenmodes.

[1074] As shown in equation (7a), each eigenmode of sufficient quality is allocated transmit power of $\left(P_{set} - \frac{1}{\text{SNR}_i}\right)$. Thus, eigenmodes that achieve better SNRs are allocated more transmit powers. The constant P_{set} determines the amounts of transmit power to allocate to the better eigenmodes. This then indirectly determines which eigenmodes get selected for use since the total available transmit power is limited and the power allocation is constrained by equation (7b).

[1075] Water-pouring analysis unit 234 thus receives the set of diagonal matrices, $\underline{\underline{\Lambda}}$, and the received noise power, σ^2. The matrices $\underline{\underline{\Lambda}}$ are then used in conjunction with the received noise power to derive a vector of scaling values, $\underline{\mathbf{b}} = [b_0 \ \ldots \ b_i \ \ldots \ b_{N_S}]^T$, where $b_i = P_i$ for $i = (1,\ 2,\ \ldots\ , N_S)$. The P_i are the solutions to the water-pouring equations (7a) and (7b). The scaling values in $\underline{\mathbf{b}}$ are indicative of the transmit powers allocated to the N_S eigenmodes, where zero or more eigenmodes may be allocated no transmit power.

[1076] **FIG. 4** is a flow diagram of an embodiment of a process 400 for allocating the total available transmit power to a set of eigenmodes. Process 400, which is one specific water-pouring implementation, determines the transmit powers, P_i, for $i \in I$, to be allocated to the eigenmodes in set I, given the total transmit power, P_{total}, available at the transmitter, the set of eigenmode total powers, $P_{i,\lambda}$, and the received noise power, σ^2.

[1077] Initially, the variable n used to denote the iteration number is set to one (i.e., $n = 1$) (step 412). For the first iteration, set $I(n)$ is defined to include all of the eigenmodes for the MIMO channel, or $I(n) = \{1,\ 2,\ \ldots\ , N_S\}$ (step 414). The cardinality (or length) of set $I(n)$ for the current iteration n is then determined as $L_I(n) = |I(n)|$, which is $L_I(n) = N_S$ for the first iteration (step 416).

[1078] The total effective power, $P_{eff}(n)$, to be distributed across the eigenmodes in set $I(n)$ is next determined (step 418). The total effective power is defined to be equal

to the total available transmit power, P_{total}, plus the sum of the inverse SNRs for the eigenmodes in set $I(n)$. This may be expressed as:

$$P_{eff}(n) = P_{total} + \sum_{i \in I(n)} \frac{\sigma^2}{P_{i,\lambda}} \quad . \qquad \text{Eq (8)}$$

[1079] The total available transmit power is then allocated to the eigenmodes in set $I(n)$. The index i used to iterate through the eigenmodes in set $I(n)$ is initialized to one (i.e., $i = 1$) (step 420). The amount of transmit power to allocate to eigenmode i is then determined (step 422) based on the following:

$$P_i(n) = \frac{P_{eff}(n)}{L_I(n)} - \frac{\sigma^2}{P_{i,\lambda}} \quad . \qquad \text{Eq (9)}$$

Each eigenmode in set $I(n)$ is allocated transmit power, P_i, in step 422. Steps 424 and 426 are part of a loop to allocate transmit power to each of the eigenmodes in set $I(n)$.

[1080] **FIG. 5A** graphically illustrates the total effective power, P_{eff}, for an example MIMO system with three eigenmodes. Each eigenmode has an inverse SNR equal to $\sigma^2 / \lambda_{ii}^2$, for $i = \{1, 2, 3\}$, which assumes a normalized transmit power of 1.0. The total transmit power available at the transmitter is $P_{total} = P_1 + P_2 + P_3$, and is represented by the shaded area in FIG. 5A. The total effective power is represented by the area in the shaded and unshaded regions in FIG. 5A.

[1081] For water-pouring, although the bottom of the vessel has an irregular surface, the water level at the top remains constant across the vessel. Likewise and as shown in FIG. 5A, after the total available transmit power, P_{total}, has been distributed to the eigenmodes, the final power level is constant across all eigenmodes. This final power level is determined by dividing $P_{eff}(n)$ by the number of eigenmodes in set $I(n)$, $L_I(n)$. The amount of power allocated to eigenmode i is then determined by subtracting the inverse SNR of that eigenmode, $\sigma^2 / \lambda_{ii}^2$, from the final power level, $P_{eff}(n) / L_I(n)$, as given by equation (9) and shown in FIG. 5A.

[1082] **FIG. 5B** shows a situation whereby the water-pouring power allocation results in an eigenmode receiving negative power. This occurs when the inverse SNR

of the eigenmode is above the final power level, which is expressed by the condition $(P_{eff}(n)/L_I(n)) < (\sigma^2/\lambda_{ii}^2)$.

[1083] Referring back to FIG. 4, at the end of the power allocation, a determination is made whether or not any eigenmode has been allocated negative power (i.e., $P_i < 0$) (step 428). If the answer is yes, then the process continues by removing from set $I(n)$ all eigenmodes that have been allocated negative powers (step 430). The index n is incremented by one (i.e., $n = n+1$) (step 432). The process then returns to step 416 to allocate the total available transmit power among the remaining eigenmodes in set $I(n)$. The process continues until all eigenmodes in set $I(n)$ have been allocated positive transmit powers, as determined in step 428. The eigenmodes not in set $I(n)$ are allocated zero power.

[1084] Water-pouring is also described by Robert G. Gallager, in "Information Theory and Reliable Communication," John Wiley and Sons, 1968, which is incorporated herein by reference. A specific algorithm for performing the basic water-pouring process for a MIMO-OFDM system is described in U.S. Patent Application Serial No. 09/978,337, entitled "Method and Apparatus for Determining Power Allocation in a MIMO Communication System," filed October 15, 2001. Water-pouring is also described in further detail in U.S. Patent Application Serial No. 10/056,275, entitled "Reallocation of Excess Power for Full Channel-State Information (CSI) Multiple-Input, Multiple-Output (MIMO) Systems," filed January 23, 2002. These applications are assigned to the assignee of the present application and incorporated herein by reference.

[1085] If water-pouring is performed to allocate the total available transmit power to the eigenmodes, then water-pouring analysis unit 234 provides a set of N_S scaling values, $\underline{\mathbf{b}} = \{b_0 \; ... \; b_i \; ... \; b_{N_S}\}$, for the N_S eigenmodes. Each scaling value is for a respective eigenmode and is used to scale the set of weights determined for that eigenmode.

[1086] For eigenmode i, a set of weights, $\hat{\underline{\mathbf{w}}}_{ii} = [\hat{w}_{ii}(0) \; ... \; \hat{w}_{ii}(k) \; ... \; \hat{w}_{ii}(N_F-1)]^T$, used to invert the channel and scale the transmit power of the eigenmode may be derived as:

$$\hat{w}_{ii}(k) = \frac{a_i b_i}{\lambda_{ii}^2(k)} \ , \qquad \text{for } k = (0,\ 1,\ \ldots,\ N_F - 1)\,, \qquad \text{Eq (10)}$$

where the normalization factor, a_i, and the scaling value, b_i, are derived as described above.

[1087] Weight computation unit 230 provides the set of weighting matrices, $\underline{\underline{\mathbf{W}}}$, which may be obtained using the weights $w_{ii}(k)$ or $\hat{w}_{ii}(k)$. Each weighting matrix, $\underline{\mathbf{W}}(k)$, is a diagonal matrix whose diagonal elements are composed of the weights derived above. In particular, if only channel inversion is performed, then each weighting matrix, $\underline{\mathbf{W}}(k)$, for $k = (0,\ 1,\ \ldots,\ N_F - 1)$, is defined as:

$$\underline{\mathbf{W}}(k) = \text{diag}\,(w_{11}(k),\ w_{22}(k),\ \ldots,\ w_{N_S N_S}(k)) \ , \qquad \text{Eq (11a)}$$

where $w_{ii}(k)$ is derived as shown in equation (3). And if both channel inversion and water-pouring are performed, then each weighting matrix, $\underline{\mathbf{W}}(k)$, for $k = (0,\ 1,\ \ldots,\ N_F - 1)$, is defined as:

$$\underline{\mathbf{W}}(k) = \text{diag}\,(\hat{w}_{11}(k),\ \hat{w}_{22}(k),\ \ldots,\ \hat{w}_{N_S N_S}(k)) \ , \qquad \text{Eq (11b)}$$

where $\hat{w}_{ii}(k)$ is derived as shown in equation (10).

[1088] Referring back to FIG. 2, a scaler/IFFT 236 receives (1) the set of unitary matrices, $\underline{\underline{\mathbf{V}}}$, which are the matrices of right eigen-vectors of $\underline{\underline{\hat{\mathbf{H}}}}$, and (2) the set of weighting matrices, $\underline{\underline{\mathbf{W}}}$, for all N_F frequency bins. Scaler/IFFT 236 then derives a spatio-temporal pulse-shaping matrix, $\underline{\mathbf{P}}_{tx}(n)$, for the transmitter based on the received matrices. Initially, the square root of each weighting matrix, $\underline{\mathbf{W}}(k)$, is computed to obtain a corresponding matrix, $\sqrt{\underline{\mathbf{W}}(k)}$, whose elements are the square roots of the elements of $\underline{\mathbf{W}}(k)$. The elements of the weighting matrices, $\underline{\mathbf{W}}(k)$ for $k = (0,\ 1,\ \ldots,\ N_F - 1)$, are related to the power of the eigenmodes. The square root then transforms the power to the equivalent signal scaling. For each frequency bin k, the product of the square-root weighting matrix, $\sqrt{\underline{\mathbf{W}}(k)}$, and the corresponding unitary

matrix, $\underline{\mathbf{V}}(k)$, is then computed to provide a product matrix, $\underline{\mathbf{V}}(k)\sqrt{\underline{\mathbf{W}}(k)}$. The set of product matrices, $\underline{\mathbf{V}}(k)\sqrt{\underline{\mathbf{W}}(k)}$ for $k = (0, 1, \ldots, N_F - 1)$, which is also denoted as $\underline{\underline{\mathbf{V}}}\sqrt{\underline{\underline{\mathbf{W}}}}$, defines the optimal or near-optimal spatio-spectral shaping to be applied to the modulation symbol vectors, $\underline{\mathbf{s}}(n)$.

[1089] An inverse FFT of $\underline{\underline{\mathbf{V}}}\sqrt{\underline{\underline{\mathbf{W}}}}$ is then computed to derive the spatio-temporal pulse-shaping matrix, $\underline{\mathbf{P}}_{tx}(n)$, for the transmitter, which may be expressed as:

$$\underline{\mathbf{P}}_{tx}(n) = \text{IFFT}\,[\underline{\underline{\mathbf{V}}}\sqrt{\underline{\underline{\mathbf{W}}}}]\ . \qquad \text{Eq (12)}$$

The pulse-shaping matrix, $\underline{\mathbf{P}}_{tx}(n)$, is an $N_T \times N_T$ matrix. Each element of $\underline{\mathbf{P}}_{tx}(n)$ is a set of N_F time-domain values, which is obtained by the inverse FFT of a set of values for the corresponding element of the product matrices, $\underline{\underline{\mathbf{V}}}\sqrt{\underline{\underline{\mathbf{W}}}}$. Each column of $\underline{\mathbf{P}}_{tx}(n)$ is a steering vector for a corresponding element of $\underline{\mathbf{s}}(n)$.

[1090] A convolver 240 receives and preconditions the modulation symbol vectors, $\underline{\mathbf{s}}(n)$, with the pulse-shaping matrix, $\underline{\mathbf{P}}_{tx}(n)$, to provide the transmitted symbol vectors, $\underline{\mathbf{x}}(n)$. In the time domain, the preconditioning is a convolution operation, and the convolution of $\underline{\mathbf{s}}(n)$ with $\underline{\mathbf{P}}_{tx}(n)$ may be expressed as:

$$\underline{\mathbf{x}}(n) = \sum_{\ell} \underline{\mathbf{P}}_{tx}(\ell)\underline{\mathbf{s}}(n-\ell)\ . \qquad \text{Eq (13)}$$

The matrix convolution shown in equation (13) may be performed as follows. To derive the i-th element of the vector $\underline{\mathbf{x}}(n)$ for time n, $x_i(n)$, the inner product of the i-th row of the matrix $\underline{\mathbf{P}}_{tx}(\ell)$ with the vector $\underline{\mathbf{s}}(n-\ell)$ is formed for a number of delay indices (e.g., $0 \le \ell \le (N_F - 1)$), and the results are accumulated to derive the element $x_i(n)$. The preconditioned symbol streams transmitted on each transmit antenna (i.e., each element of $\underline{\mathbf{x}}(n)$ or $x_i(n)$) is thus formed as a weighted combination of the N_R modulation symbol streams, with the weighting determined by the appropriate column of the matrix $\underline{\mathbf{P}}_{tx}(n)$. The process is repeated such that each element of $\underline{\mathbf{x}}(n)$ is obtained from a respective column of the matrix $\underline{\mathbf{P}}_{tx}(n)$ and the vector $\underline{\mathbf{s}}(n)$.

[1091] Each element of $\underline{\mathbf{x}}(n)$ corresponds to a sequence of preconditioned symbols to be transmitted over a respective transmit antenna. The N_T preconditioned symbol sequences collectively form a sequence of vectors, which are also referred to as the transmitted symbol vectors, $\underline{\mathbf{x}}(n)$, with each such vector including up to N_T preconditioned symbols to be transmitted from up to N_T transmit antennas for the n-th symbol period. The N_T preconditioned symbol sequences are provided to transmitters 122a through 122t and processed to derive N_T modulated signals, which are then transmitted from antennas 124a through 124t, respectively.

[1092] The embodiment shown in FIG. 2 performs time-domain beam-steering of the modulation symbol vectors, $\underline{\mathbf{s}}(n)$. The beam-steering may also be performed in the frequency domain. This can be done using techniques, such as the overlap-add method, which are well-known in the digital signal processing field, for implementing finite-duration impulse response (FIR) filters in the frequency domain. In this case, the sequences that make up the elements of the matrix $\underline{\mathbf{P}}_{tx}(n)$ for $n=(0,\ 1,\ \ldots\ ,N_F-1)$ are each padded with N_O-N_F zeros, resulting in a matrix of zero-padded sequences, $\underline{\mathbf{q}}_{tx}(n)$, for $n=(0,\ 1,\ \ldots\ ,N_O-1)$. An N_O-point fast Fourier transform (FFT) is then computed for each zero-padded sequence in the matrix $\underline{\mathbf{q}}_{tx}(n)$, resulting in a matrix $\underline{\mathbf{Q}}_{tx}(k)$ for $k=(0,\ 1,\ \ldots\ ,N_O-1)$.

[1093] Also, the sequences of modulation symbols that make up the elements of $\underline{\mathbf{s}}(n)$ are each broken up into subsequences of length $N_{SS}=N_O-N_F+1$. Each subsequence is then padded with N_F-1 zeros to provide a corresponding vector of length N_O. The sequences of $\underline{\mathbf{s}}(n)$ are thus processed to provide sequences of vectors of length N_O, $\tilde{\underline{\mathbf{s}}}_\ell(n)$, where the subscript ℓ is the index for the vectors that correspond to the zero-padded subsequences. An N_O-point fast Fourier transform is then computed for each of the zero-padded subsequences, resulting in a sequence of frequency-domain vectors, $\tilde{\underline{\mathbf{S}}}_\ell(k)$, for different values of ℓ. Each vector $\tilde{\underline{\mathbf{S}}}_\ell(k)$, for a given ℓ, includes a set of frequency-domain vectors of length N_O (i.e., for $k=(0,\ 1,\ \ldots\ ,N_O-1)$). The matrix $\underline{\mathbf{Q}}_{tx}(k)$ is then multiplied with the vector $\tilde{\underline{\mathbf{S}}}_\ell(k)$, for each value of ℓ, where the pre-multiplication is performed for each value of k, i.e., for $k=(0,\ 1,\ \ldots\ ,N_O-1)$.

The inverse FFTs are then computed for the matrix-vector product $\underline{\mathbf{Q}}_{tx}(k)\tilde{\underline{\mathbf{S}}}_{\ell}(k)$ to provide a set of time-domain subsequences of length N_O. The resulting subsequences are then reassembled, according to the overlap-add method, or similar means, as is well-known in the art, to form the desired output sequences.

[1094] **FIG. 6** is a flow diagram of an embodiment of a process 600 that may be performed at the transmitter unit to implement the various transmit processing techniques described herein. Initially, data to be transmitted (i.e., the information bits) is processed in accordance with a particular processing scheme to provide a number of streams of modulation symbols (step 612). As noted above, the processing scheme may include one or more coding schemes and one or more modulation schemes (e.g., a separate coding and modulation scheme for each modulation symbol stream).

[1095] An estimated channel response matrix for the MIMO channel is then obtained (step 614). This matrix may be the estimated channel impulse response matrix, $\underline{\underline{\hat{\mathcal{H}}}}$, or the estimated channel frequency response matrix, $\underline{\underline{\hat{\mathbf{H}}}}$, which may be provided to the transmitter from the receiver. The estimated channel response matrix is then decomposed (e.g., using channel eigen-decomposition) to obtain a set of matrices of right eigen-vectors, $\underline{\underline{\mathbf{V}}}$, and a set of matrices of singular values, $\underline{\underline{\Lambda}}$ (step 616).

[1096] A number of sets of weights, $\underline{\mathbf{w}}_{ii}$, are then derived based on the matrices of singular values (step 618). One set of weight may be derived for each eigenmode used for data transmission. These weights are used to reduce or minimize intersymbol interference at the receiver by inverting the frequency response of each eigenmode selected for use.

[1097] A set of scaling values, $\underline{\mathbf{b}}$, may also be derived based on the matrices of singular values (step 620). Step 620 is optional, as indicated by the dashed box for step 620 in FIG. 6. The scaling values may be derived using water-pouring analysis and are used to adjust the transmit powers of the selected eigenmodes.

[1098] A pulse-shaping matrix, $\underline{\mathbf{P}}_{tx}(n)$, is then derived based on the matrices of right eigen-vectors, $\underline{\underline{\mathbf{V}}}$, the sets of weights, $\underline{\mathbf{w}}_{ii}$, and (if available) the set of scaling values, $\underline{\mathbf{b}}$ (step 622). The streams of modulation symbols are then preconditioned (in either the time domain or frequency domain) based on the pulse-shaping matrix to

provide a number of streams of preconditioned symbols, $\underline{\mathbf{x}}(n)$, to be transmitted over the MIMO channel (step 624).

[1099] Time-domain transmit processing with channel eigenmode decomposition and water-pouring is described in further detail in U.S. Patent Application Serial No. 10/017,038, entitled "Time-Domain Transmit and Receive Processing with Channel Eigen-mode Decomposition for MIMO Systems," filed December 7, 2001, which is assigned to the assignee of the present application and incorporated herein by reference.

[1100] **FIG. 7** is a block diagram of an embodiment of a receiver unit 700 capable of implementing various processing techniques described herein. Receiver unit 700 is an embodiment of the receiver portion of receiver system 150 in FIG. 1. Receiver unit 700 includes (1) a RX MIMO processor 160a that processes N_R received symbol streams to provide N_T recovered symbol streams and (2) a RX data processor 162a that demodulates, deinterleaves, and decodes the recovered symbols to provide decoded bits. RX MIMO processor 160a and RX data processor 162a are one embodiment of RX MIMO processor 160 and RX data processor 162, respectively, in FIG. 1.

[1101] Referring back to FIG. 1, the transmitted signals from N_T transmit antennas are received by each of N_R antennas 152a through 152r. The received signal from each antenna is routed to a respective receiver 154, which is also referred to as a front-end processor. Each receiver 154 conditions (e.g., filters, amplifies, and frequency downconverts) a respective received signal, and further digitizes the conditioned signal to provide ADC samples. Each receiver 154 may further data demodulate the ADC samples with a recovered pilot to provide a respective stream of received symbols. Receivers 154a through 154r thus provide N_R received symbol streams. These streams collectively form a sequence of vectors, which are also referred to as the received symbol vectors, $\underline{\mathbf{r}}(n)$, with each such vector including N_R received symbols from the N_R receivers 154 for the n-th symbol period. The received symbol vectors, $\underline{\mathbf{r}}(n)$, are then provided to RX MIMO processor 160a.

[1102] Within RX MIMO processor 160a, a channel estimator 712 receives the vectors $\underline{\mathbf{r}}(n)$ and derives an estimated channel impulse response matrix, $\hat{\underline{\underline{\mathcal{H}}}}$, which may be sent back to the transmitter system and used in the transmit processing. An FFT 714

then performs an FFT on the estimated channel impulse response matrix, $\hat{\underline{\underline{\mathcal{H}}}}$, to obtain the estimated channel frequency response matrix, $\hat{\underline{\underline{\mathbf{H}}}}$ (i.e., $\hat{\underline{\underline{\mathbf{H}}}} = \text{FFT}[\hat{\underline{\underline{\mathcal{H}}}}]$).

[1103] A unit 716 then performs the channel eigen-decomposition of $\hat{\underline{\mathbf{H}}}(k)$, for each frequency bin k, to obtain the corresponding matrix of left eigen-vectors, $\underline{\mathbf{U}}(k)$. Each column of $\underline{\underline{\mathbf{U}}}$, where $\underline{\underline{\mathbf{U}}} = [\underline{\mathbf{U}}(0) \ \ldots \ \underline{\mathbf{U}}(k) \ \ldots \ \underline{\mathbf{U}}(N_F - 1)]$, is a steering vector for a corresponding element of $\underline{\mathbf{r}}(n)$, and is used to orthogonalize the received symbol streams. An IFFT 718 then performs the inverse FFT of $\underline{\underline{\mathbf{U}}}$ to obtain a spatio-temporal pulse-shaping matrix, $\underline{\mathcal{U}}(n)$, for the receiver system.

[1104] A convolver 720 then conditions the received symbol vectors, $\underline{\mathbf{r}}(n)$, with the conjugate transpose of the spatio-temporal pulse-shaping matrix, $\underline{\mathcal{U}}^H(n)$, to obtain the recovered symbol vectors, $\hat{\underline{\mathbf{s}}}(n)$, which are estimates of the modulation symbol vectors, $\underline{\mathbf{s}}(n)$. In the time domain, the conditioning is a convolution operation, which may be expressed as:

$$\hat{\underline{\mathbf{s}}}(n) = \sum_{\ell} \underline{\mathcal{U}}^H(\ell)\, \underline{\mathbf{r}}(n-\ell) \ . \qquad \text{Eq (14)}$$

[1105] The pulse-shaping at the receiver may also be performed in the frequency domain, similar to that described above for the transmitter. In this case, the N_R sequences of received symbols for the N_R receive antennas, which make up the sequence of received symbol vectors, $\underline{\mathbf{r}}(n)$, are each broken up into subsequences of N_{SS} received symbols, and each subsequence is zero-padded to provide a corresponding vector of length N_O. The N_R sequences of $\underline{\mathbf{r}}(n)$ are thus processed to provide N_R sequences of vectors of length N_O, $\tilde{\underline{\mathbf{r}}}_\ell(n)$, where the subscript ℓ is the index for the vectors that correspond to the zero-padded subsequences. Each zero-padded subsequence is then transformed via an FFT, resulting in a sequence of frequency-domain vectors, $\underline{\mathbf{R}}_\ell(k)$, for different values of ℓ. Each vector $\underline{\mathbf{R}}_\ell(k)$, for a given ℓ, includes a set of frequency-domain vectors of length N_O (i.e., for $k = (0,\ 1,\ \ldots\ , N_O - 1)$).

[1106] The conjugate transpose of the spatio-temporal pulse-shaping matrix, $\underline{\mathcal{U}}^H(n)$, is also zero-padded and transformed via an FFT to obtain a frequency-domain

matrix, $\underline{\tilde{\mathbf{U}}}^H(k)$ for $k = (0,\ 1,\ \ldots\ , N_O - 1)$. The vector $\underline{\mathbf{R}}_\ell(k)$, for each value of ℓ, is then pre-multiplied with the conjugate transpose matrix $\underline{\tilde{\mathbf{U}}}^H(k)$ (where the pre-multiplication is performed for each value of k, i.e., for $k = (0,\ 1,\ \ldots\ , N_O - 1)$) to obtain a corresponding frequency-domain vector $\underline{\hat{\mathbf{S}}}_\ell(k)$. Each vector $\underline{\hat{\mathbf{S}}}_\ell(k)$, which includes a set of frequency-domain vectors of length N_O, can then be transformed via an inverse FFT to provide a corresponding set of time-domain subsequences of length N_O. The resulting subsequences are then reassembled according to the overlap-add method, or similar means, as is well-known in the art, to obtain sequences of recovered symbols, which corresponds to the set of recovered symbol vectors, $\underline{\hat{\mathbf{s}}}(n)$.

[1107] Thus recovered symbol vectors, $\underline{\hat{\mathbf{s}}}(n)$, may be characterized as a convolution in the time domain, as follows:

$$\underline{\hat{\mathbf{s}}}(n) = \sum_\ell \underline{\Gamma}(\ell)\,\underline{\mathbf{s}}(n-\ell) + \underline{\hat{\mathbf{z}}}(n) \quad , \qquad \text{Eq (15)}$$

where $\underline{\Gamma}(\ell)$ is the inverse FFT of $\underline{\hat{\Lambda}}(k) = \underline{\Lambda}(k)\sqrt{\underline{\mathbf{W}}(k)}$; and

$\underline{\hat{\mathbf{z}}}(n)$ is the received noise as transformed by the receiver spatio-temporal pulse-shaping matrix, $\underline{\mathcal{U}}^H(\ell)$.

The matrix $\underline{\Gamma}(n)$ is a diagonal matrix of eigen-pulses derived from the set of matrices $\underline{\underline{\hat{\Lambda}}}$, where $\underline{\underline{\hat{\Lambda}}} = [\underline{\hat{\Lambda}}(0)\ \ldots\ \underline{\hat{\Lambda}}(k)\ \ldots\ \underline{\hat{\Lambda}}(N_F - 1)]$. In particular, each diagonal element of $\underline{\Gamma}(n)$ corresponds to an eigen-pulse that is obtained as the IFFT of a set of singular values, $[\hat{\lambda}_{ii}(0)\ \ldots\ \hat{\lambda}_{ii}(k)\ \ldots\ \hat{\lambda}_{ii}(N_F - 1)]^T$, for a corresponding element of $\underline{\underline{\hat{\Lambda}}}$,

[1108] The two forms for ordering the singular values, sorted and random-ordered, result in two different types of eigen-pulses. For the sorted form, the resulting eigen-pulse matrix, $\underline{\Gamma}_s(n)$, is a diagonal matrix of pulses that are sorted in descending order of energy content. The pulse corresponding to the first diagonal element of the eigen-pulse matrix, $\{\underline{\Gamma}_s(n)\}_{11}$, has the most energy, and the pulses corresponding to elements further down the diagonal have successively less energy. Furthermore, when the SNR is low enough that water-pouring results in some of the frequency bins having little or no energy, the energy is taken out of the smallest eigen-pulses first. Thus, at low SNRs,

one or more of the eigen-pulses may have little or no energy. This has the advantage that at low SNRs, the coding and modulation are simplified through the reduction in the number of orthogonal subchannels. However, in order to approach channel capacity, the coding and modulation are performed separately for each eigen-pulse.

[1109] The random-ordered form of the singular values in the frequency domain may be used to further simplify coding and modulation (i.e., to avoid the complexity of separate coding and modulation for each element of the eigen-pulse matrix). In the random-ordered form, for each frequency bin, the ordering of the singular values is random instead of being based on their magnitude or size. This random ordering can result in approximately equal energy in all of the eigen-pulses. When the SNR is low enough to result in frequency bins with little or no energy, these bins are spread approximately evenly among the eigenmodes so that the number of eigen-pulses with non-zero energy is the same independent of SNR. At high SNRs, the random-order form has the advantage that all of the eigen-pulses have approximately equal energy, in which case separate coding and modulation for different eigenmodes are not required.

[1110] If the response of the MIMO channel is frequency selective, then the elements in the diagonal matrices, $\underline{\Lambda}(k)$, are time-dispersive. However, because of the pre-processing at the transmitter to invert the channel, the resulting recovered symbol sequences, $\hat{\underline{s}}(n)$, have little intersymbol interference, if the channel inversion is effectively performed. In that case, additional equalization would not be required at the receiver to achieve high performance.

[1111] If the channel inversion is not effective (e.g., due to an inaccurate estimated channel frequency response matrix, $\hat{\underline{\underline{H}}}$) then an equalizer may be used to equalize the recovered symbols, $\hat{\underline{s}}(n)$, prior to the demodulation and decoding. Various types of equalizer may be used to equalize the recovered symbol streams, including a minimum mean square error linear equalizer (MMSE-LE), a decision feedback equalizer (DFE), a maximum likelihood sequence estimator (MLSE), and so on.

[1112] Since the orthogonalization process at the transmitter and receiver results in decoupled (i.e., orthogonal) recovered symbol streams at the receiver, the complexity of the equalization required for the decoupled symbol streams is greatly reduced. In particular, the equalization may be achieved by parallel time-domain equalization of the independent symbol streams. The equalization may be performed as described in the

aforementioned U.S. Patent Application Serial No. 10/017,038, and in U.S. Patent Application Serial No. 09/993,087, entitled "Multiple-Access Multiple-Input Multiple-Output (MIMO) Communication System," filed November 6, 2001, which is assigned to the assignee of the present application and incorporated herein by reference.

[1113] For the embodiment in FIG. 7, the recovered symbol vectors, $\hat{\underline{s}}(n)$, are provided to RX data processor 162a. Within processor 162a, a symbol demapping element 732 demodulates each recovered symbol in $\hat{\underline{s}}(n)$ in accordance with a demodulation scheme that is complementary to the modulation scheme used for that symbol at the transmitter system. The demodulated data from symbol demapping element 732 is then de-interleaved by a deinterleaver 734. The deinterleaved data is further decoded by a decoder 736 to obtain the decoded bits, $\hat{d}_i$, which are estimates of the transmitted information bits, d_i. The deinterleaving and decoding are performed in a manner complementary to the interleaving and encoding, respectively, performed at the transmitter system. For example, a Turbo decoder or a Viterbi decoder may be used for decoder 736 if Turbo or convolutional coding, respectively, is performed at the transmitter system.

[1114] **FIG. 8** is a flow diagram of a process 800 that may be performed at the receiver unit to implement the various receive processing techniques described herein. Initially, an estimated channel response matrix for the MIMO channel is obtained (step 812). This matrix may be the estimated channel impulse response matrix, $\hat{\underline{\underline{\mathcal{H}}}}$, or the estimated channel frequency response matrix, $\hat{\underline{\underline{\mathbf{H}}}}$. The matrix $\hat{\underline{\underline{\mathcal{H}}}}$ or $\hat{\underline{\underline{\mathbf{H}}}}$ may be obtained, for example, based on pilot symbols transmitted over the MIMO channel. The estimated channel response matrix is then decomposed (e.g., using channel eigen-decomposition) to obtain a set of matrices of left eigen-vectors, $\underline{\underline{\mathbf{U}}}$ (step 814).

[1115] A pulse-shaping matrix $\underline{\mathcal{U}}(n)$ is then derived based on the matrices of left eigen-vectors, $\underline{\underline{\mathbf{U}}}$ (step 816). The streams of received symbols are then conditioned (in either the time domain or frequency domain) based on the pulse-shaping matrix $\underline{\mathcal{U}}(n)$ to provide the streams of recovered symbols (step 818). The recovered symbols are further processed in accordance with a particular receive processing scheme, which is

complementary to the transmit processing scheme used at the transmitter, to provide the decoded data (step 820).

[1116] Time-domain receive processing with channel eigenmode decomposition is described in further detail in the aforementioned U.S. Patent Application Serial No. 10/017,038.

[1117] The techniques for processing a data transmission at a transmitter and a receiver described herein may be implemented in various wireless communication systems, including but not limited to MIMO and CDMA systems. These techniques may also be used for the forward link and/or the reverse link.

[1118] The techniques described herein to process a data transmission at the transmitter and receiver may be implemented by various means. For example, these techniques may be implemented in hardware, software, or a combination thereof. For a hardware implementation, the elements used to perform various signal processing steps at the transmitter (e.g., to code and modulate the data, decompose the channel response matrix, derive the weights to invert the channel, derive the scaling values for power allocation, derive the transmitter pulse-shaping matrix, precondition the modulation symbols, and so on) or at the receiver (e.g., to decompose the channel response matrix, derive the receiver pulse-shaping matrix, condition the received symbols, demodulate and decode the recovered symbols, and so on) may be implemented within one or more application specific integrated circuits (ASICs), digital signal processors (DSPs), digital signal processing devices (DSPDs), programmable logic devices (PLDs), field programmable gate arrays (FPGAs), processors, controllers, micro-controllers, microprocessors, other electronic units designed to perform the functions described herein, or a combination thereof.

[1119] For a software implementation, some or all of the signal processing steps at each of the transmitter and receiver may be implemented with modules (e.g., procedures, functions, and so on) that perform the functions described herein. The software codes may be stored in a memory unit (e.g., memories 132 and 172 in FIG. 1) and executed by a processor (e.g., controllers 130 and 170). The memory unit may be implemented within the processor or external to the processor, in which case it can be communicatively coupled to the processor via various means as is known in the art.

[1120] The previous description of the disclosed embodiments is provided to enable any person skilled in the art to make or use the present invention. Various

modifications to these embodiments will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other embodiments without departing from the spirit or scope of the invention. Thus, the present invention is not intended to be limited to the embodiments shown herein but is to be accorded the widest scope consistent with the principles and novel features disclosed herein.

[1121] **WHAT IS CLAIMED IS:**

CLAIMS

1. In a multiple-input multiple-output (MIMO) communication system, a method for processing data for transmission over a MIMO channel, comprising:

processing data in accordance with a particular processing scheme to provide a plurality of streams of modulation symbols;

deriving a pulse-shaping matrix based on an estimated response of the MIMO channel and in a manner to reduce intersymbol interference at a receiver; and

preconditioning the plurality of modulation symbol streams based on the pulse-shaping matrix to provide a plurality of streams of preconditioned symbols for transmission over the MIMO channel.

2. The method of claim 1, further comprising:

deriving a plurality of weights based on an estimated channel response matrix for the MIMO channel, wherein the weights are used to invert a frequency response of the MIMO channel, and wherein the pulse-shaping matrix is further derived based on the weights.

3. The method of claim 2, further comprising:

decomposing the estimated channel response matrix to obtain a plurality of matrices of eigen-vectors and a plurality of matrices of singular values, and

wherein the weights are derived based on the matrices of singular values and the pulse-shaping matrix is further derived based on the matrices of eigen-vectors.

4. The method of claim 2, wherein the estimated channel response matrix is descriptive of a plurality of eigenmodes of the MIMO channel.

5. The method of claim 4, wherein one set of weights is derived for each eigenmode used for data transmission and wherein the weights in each set are derived to invert the frequency response of the corresponding eigenmode.

6. The method of claim 4, further comprising:

deriving a plurality of scaling values based on the matrices of singular values, wherein the scaling values are used to adjust transmit powers for the eigenmodes of the MIMO channel, and wherein the pulse-shaping matrix is further derived based on the scaling values.

7. The method of claim 6, wherein the scaling values are derived based on water-pouring analysis.

8. The method of claim 3, wherein the estimated channel response matrix is provided in the frequency domain and is decomposed in the frequency domain.

9. The method of claim 3, wherein the estimated channel response matrix is decomposed using channel eigen-decomposition.

10. The method of claim 4, wherein eigenmodes associated with transmission capabilities below a particular threshold are not used for data transmission.

11. The method of claim 3, wherein the singular values in each matrix are sorted based on their magnitude.

12. The method of claim 4, wherein the singular values in each matrix are randomly ordered such that the eigenmodes of the MIMO channel are associated with approximately equal transmission capabilities.

13. The method of claim 1, wherein the pulse-shaping matrix comprises a plurality of sequences of time-domain values, and wherein the preconditioning is performed in the time domain by convolving the streams of modulation symbols with the pulse-shaping matrix.

14. The method of claim 1, wherein the pulse-shaping matrix comprises a plurality of sequences of frequency-domain values, and wherein the preconditioning is performed in the frequency domain by multiplying a plurality of streams of transformed modulation symbols with the pulse-shaping matrix.

15. The method of claim 1, wherein the pulse-shaping matrix is derived to maximize capacity by allocating more transmit power to eigenmodes of the MIMO channel having greater transmission capabilities.

16. The method of claim 1, wherein the pulse-shaping matrix is derived to provide approximately equal received signal-to-noise-and-interference ratios (SNRs) for the plurality of modulation symbol streams at the receiver.

17. The method of claim 1, wherein the particular processing scheme defines a separate coding and modulation scheme for each modulation symbol stream.

18. The method of claim 1, wherein the particular processing scheme defines a common coding and modulation scheme for all modulation symbol streams.

19. In a multiple-input multiple-output (MIMO) communication system, a method for processing data for transmission over a MIMO channel, comprising:

processing data in accordance with a particular processing scheme to provide a plurality of streams of modulation symbols;

obtaining an estimated channel response matrix for the MIMO channel;

decomposing the estimated channel response matrix to obtain a plurality of matrices of eigen-vectors and a plurality of matrices of singular values;

deriving a plurality of weights based on the matrices of singular values, wherein the weights are used to invert the frequency response of the MIMO channel;

deriving a pulse-shaping matrix based on the matrices of eigen-vectors and the weights; and

preconditioning the plurality of streams of modulation symbols based on the pulse-shaping matrix to provide a plurality of streams of preconditioned symbols for transmission over the MIMO channel.

20. The method of claim 19, further comprising:

deriving a plurality of scaling values based on the matrices of singular values, wherein the scaling values are used to adjust transmit powers for eigenmodes of the

MIMO channel, and wherein the pulse-shaping matrix is further derived based on the scaling values.

21. A memory communicatively coupled to a digital signal processing device (DSPD) capable of interpreting digital information to:

process data in accordance with a particular processing scheme to provide a plurality of streams of modulation symbols;

derive a pulse-shaping matrix based on an estimated response of the MIMO channel and in a manner to reduce intersymbol interference at a receiver; and

precondition the plurality of streams of modulation symbols based on the pulse-shaping matrix to provide a plurality of streams of preconditioned symbols for transmission over the MIMO channel.

22. In a multiple-input multiple-output (MIMO) communication system, a method for processing a data transmission received via a MIMO channel, comprising:

obtaining an estimated channel response matrix for the MIMO channel;

decomposing the estimated channel response matrix to obtain a plurality of matrices of eigen-vectors;

deriving a pulse-shaping matrix based on the matrices of eigen-vectors; and

conditioning a plurality of streams of received symbols based on the pulse-shaping matrix to provide a plurality of streams of recovered symbols which are estimates of modulation symbols transmitted for the data transmission, wherein the modulation symbols are preconditioned at a transmitter, prior to transmission over the MIMO channel, in a manner to reduce intersymbol interference at a receiver.

23. The method of claim 22, wherein the conditioning is performed in the time domain based on a time-domain pulse-shaping matrix.

24. The method of claim 22, wherein the conditioning is performed in the frequency domain and includes

transforming the plurality of received symbol streams to the frequency domain;

multiplying the transformed received symbol streams with a frequency-domain pulse-shaping matrix to provide a plurality of conditioned symbol streams; and

transforming the plurality of conditioned symbol streams to the time domain to provide the plurality of recovered symbol streams.

25. The method of claim 22, wherein the conditioning orthogonalizes a plurality of streams of modulation symbols transmitted over the MIMO channel.

26. The method of claim 22, further comprising:

demodulating the plurality of recovered symbol streams in accordance with one or more demodulation schemes to provide a plurality of demodulated data streams; and

decoding the plurality of demodulated data streams in accordance with one or more decoding schemes to provide decoded data.

27. The method of claim 22, further comprising:

deriving channel state information (CSI) comprised of the estimated channel response matrix for the MIMO channel; and

sending the CSI back to the transmitter.

28. In a multiple-input multiple-output (MIMO) communication system, a method for processing a data transmission received via a MIMO channel, comprising:

obtaining an estimated channel response matrix for the MIMO channel;

decomposing the estimated channel response matrix to obtain a plurality of matrices of eigen-vectors;

deriving a pulse-shaping matrix based on the matrices of eigen-vectors;

conditioning a plurality of streams of received symbols based on the pulse-shaping matrix to provide a plurality of streams of recovered symbols which are estimates of modulation symbols transmitted for the data transmission, wherein the modulation symbols are preconditioned at a transmitter, prior to transmission over the MIMO channel, in a manner to reduce intersymbol interference at a receiver;

demodulating the plurality of recovered symbol streams in accordance with one or more demodulation schemes to provide a plurality of demodulated data streams; and

decoding the plurality of demodulated data streams in accordance with one or more decoding schemes to provide decoded data.

29. A memory communicatively coupled to a digital signal processing device (DSPD) capable of interpreting digital information to:

obtain an estimated channel response matrix for a MIMO channel used for a data transmission;

decompose the estimated channel response matrix to obtain a plurality of matrices of eigen-vectors;

derive a pulse-shaping matrix based on the matrices of eigen-vectors; and

condition a plurality of streams of received symbols based on the pulse-shaping matrix to provide a plurality of streams of recovered symbols which are estimates of modulation symbols transmitted for the data transmission, wherein the modulation symbols are preconditioned at a transmitter, prior to transmission over the MIMO channel, in a manner to reduce intersymbol interference at a receiver.

30. A transmitter unit in a multiple-input multiple-output (MIMO) communication system, comprising:

a TX data processor operative to process data in accordance with a particular processing scheme to provide a plurality of streams of modulation symbols; and

a TX MIMO processor operative to derive a pulse-shaping matrix based on an estimated response of a MIMO channel and in a manner to reduce intersymbol interference at a receiver, and to precondition the plurality of modulation symbol streams based on the pulse-shaping matrix to provide a plurality of streams of preconditioned symbols for transmission over the MIMO channel.

31. The transmitter unit of claim 30, wherein the TX MIMO processor is further operative to derive a plurality of weights based on an estimated channel response matrix for the MIMO channel, wherein the weights are used to invert the frequency response of the MIMO channel, and wherein the pulse-shaping matrix is derived based in part on the weights.

32. The transmitter unit of claim 31, wherein the TX MIMO processor is further operative to decompose the estimated channel response matrix to obtain a plurality of matrices of eigen-vectors and a plurality of matrices of singular values, and

wherein the weights are derived based on the matrices of singular values and the pulse-shaping matrix is further derived based on the matrices of eigen-vectors.

33. The transmitter unit of claim 31, wherein the TX MIMO processor is further operative to derive a plurality of scaling values used to adjust transmit powers for the eigenmodes of the MIMO channel, and wherein the pulse-shaping matrix is further derived based on the scaling values.

34. The transmitter unit of claim 33, wherein the scaling values are derived based on water-pouring analysis on a plurality of matrices of singular values obtained from the estimated channel response matrix.

35. An apparatus in a multiple-input multiple-output (MIMO) communication system, comprising:

means for processing data in accordance with a particular processing scheme to provide a plurality of streams of modulation symbols;

means for deriving a pulse-shaping matrix based on an estimated response of a MIMO channel and in a manner to reduce intersymbol interference at a receiver; and

means for preconditioning the plurality of modulation symbol streams based on the pulse-shaping matrix to provide a plurality of streams of preconditioned symbols for transmission over the MIMO channel.

36. A digital signal processor comprising:

means for processing data in accordance with a particular processing scheme to provide a plurality of streams of modulation symbols;

means for deriving a pulse-shaping matrix based on an estimated response of a multiple-input multiple-output (MIMO) channel and in a manner to reduce intersymbol interference at a receiver; and

means for preconditioning the plurality of modulation symbol streams based on the pulse-shaping matrix to provide a plurality of streams of preconditioned symbols for transmission over the MIMO channel.

37. A receiver unit in a multiple-input multiple-output (MIMO) communication system, comprising:

an RX MIMO processor operative to obtain an estimated channel response matrix for a MIMO channel used for a data transmission, decompose the estimated channel response matrix to obtain a plurality of matrices of eigen-vectors, derive a pulse-shaping matrix based on the matrices of eigen-vectors, and condition a plurality of streams of received symbols based on the pulse-shaping matrix to obtain a plurality of streams of recovered symbols which are estimates of modulation symbols transmitted over the MIMO channel, wherein the modulation symbols were preconditioned at a transmitter, prior to transmission over the MIMO channel, in a manner to reduce intersymbol interference at the receiver unit; and

an RX data processor operative to process the plurality of recovered symbol streams in accordance with a particular processing scheme to provide decoded data.

38. The receiver unit of claim 37, wherein the RX MIMO processor is operative to condition the plurality of streams of received symbols in the time domain based on a time-domain pulse-shaping matrix.

39. An apparatus in a multiple-input multiple-output (MIMO) communication system, comprising:

means for obtaining an estimated channel response matrix for a MIMO channel used for a data transmission;

means for decomposing the estimated channel response matrix to obtain a plurality of matrices of eigen-vectors;

means for deriving a pulse-shaping matrix based on the matrices of eigen-vectors; and

means for conditioning a plurality of streams of received symbols based on the pulse-shaping matrix to provide a plurality of streams of recovered symbols which are estimates of modulation symbols transmitted for the data transmission, wherein the modulation symbols are preconditioned at a transmitter, prior to transmission over the MIMO channel, in a manner to reduce intersymbol interference at a receiver.

40. A digital signal processor comprising:

means for obtaining an estimated channel response matrix for a multiple-input multiple-output (MIMO) channel used for a data transmission;

means for decomposing the estimated channel response matrix to obtain a plurality of matrices of eigen-vectors;

means for deriving a pulse-shaping matrix based on the matrices of eigen-vectors; and

means for conditioning a plurality of streams of received symbols based on the pulse-shaping matrix to provide a plurality of streams of recovered symbols which are estimates of modulation symbols transmitted for the data transmission, wherein the modulation symbols are preconditioned at a transmitter, prior to transmission over the MIMO channel, in a manner to reduce intersymbol interference at a receiver.

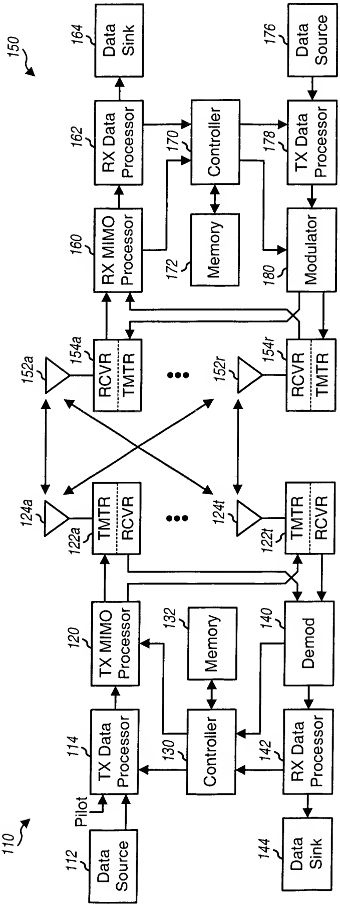

110
112 Data Source
Pilot
114 TX Data Processor
120 TX MIMO Processor
130 Controller
132 Memory
140 Demod
142 RX Data Processor
144 Data Sink
122a TMTR RCVR
124a
122t TMTR RCVR
124t
152a
154a RCVR TMTR
152r
154r RCVR TMTR
150
160 RX MIMO Processor
162 RX Data Processor
164 Data Sink
170 Controller
172 Memory
176 Data Source
178 TX Data Processor
180 Modulator

FIG. 1

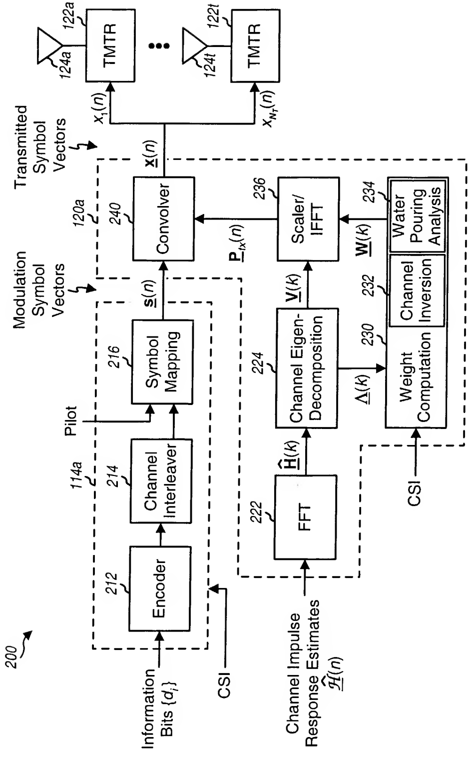

200
Information Bits $\{d_i\}$
CSI
114a
212
Encoder
214
Channel Interleaver
Pilot
216
Symbol Mapping
Modulation Symbol Vectors
$\underline{s}(n)$
120a
240
Convolver
Transmitted Symbol Vectors
$\underline{x}(n)$
$x_1(n)$
$x_{N_T}(n)$
122a
TMTR
124a
122t
TMTR
124t
Channel Impulse Response Estimates $\hat{\underline{\mathcal{H}}}(n)$
222
FFT
$\hat{\underline{H}}(k)$
224
Channel Eigen-Decomposition
$\underline{V}(k)$
$\underline{\Lambda}(k)$
236
Scaler/ IFFT
$\underline{P}_{tx}(n)$
$\underline{W}(k)$
230
Weight Computation
232
Channel Inversion
234
Water Pouring Analysis
CSI

FIG. 2

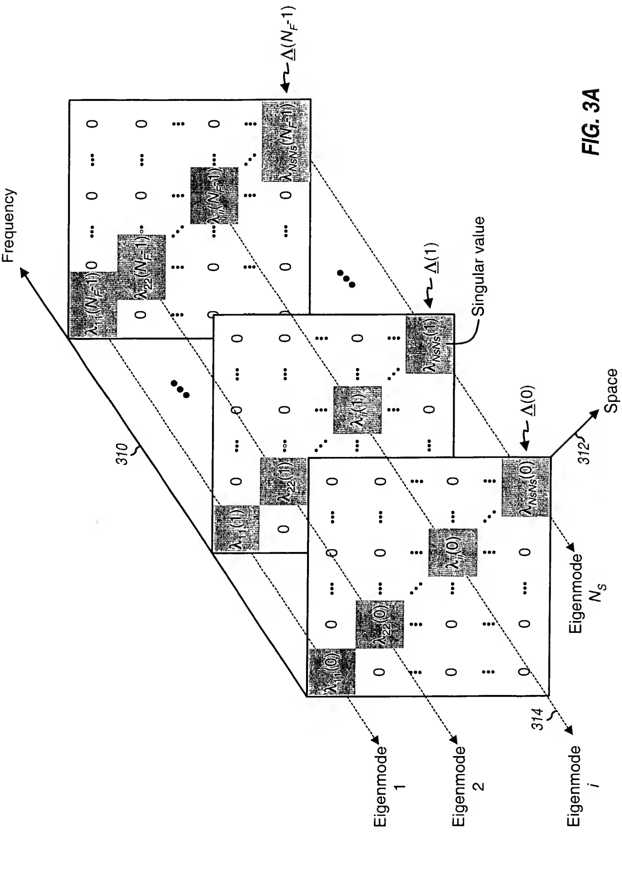

Frequency
Space
Singular value
Eigenmode 1
Eigenmode 2
Eigenmode i
Eigenmode N_S
310
312
314
$\underline{\Delta}(0)$
$\underline{\Delta}(1)$
$\underline{\Delta}(N_F-1)$

FIG. 3A

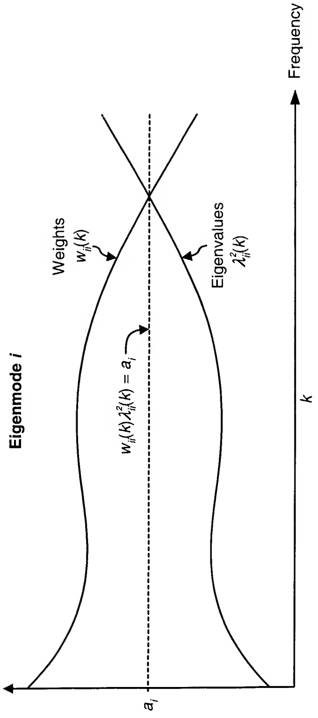

Eigenmode i
Weights $w_{ii}(k)$
Eigenvalues $\lambda^2_{ii}(k)$
$w_{ii}(k)\lambda^2_{ii}(k) = a_i$
a_i
k
Frequency

FIG. 3B

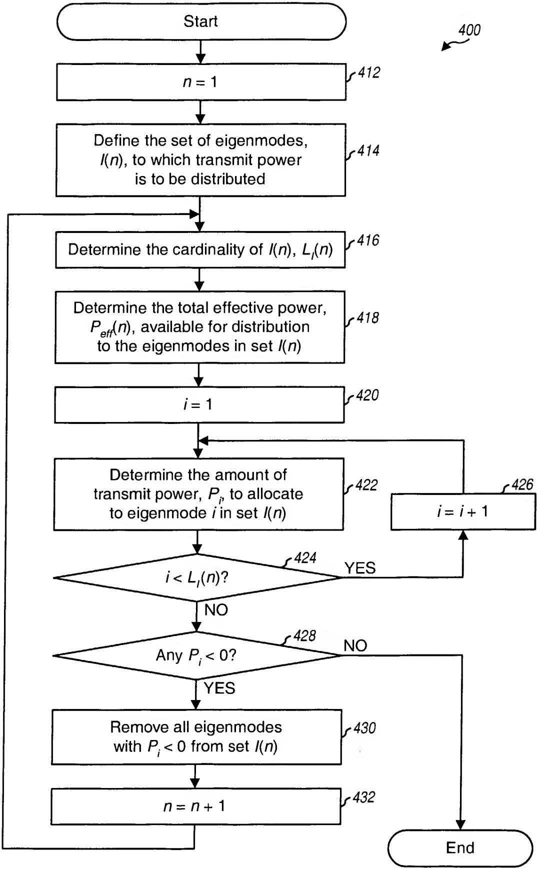
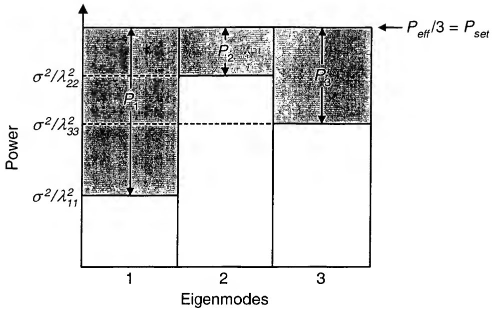
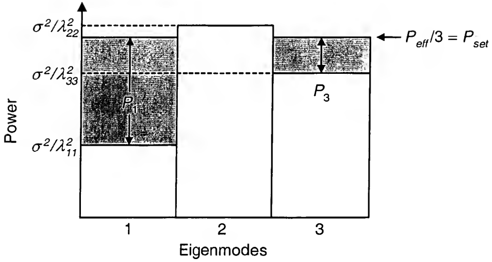

400
Start
n = 1
412
Define the set of eigenmodes, I(n), to which transmit power is to be distributed
414
Determine the cardinality of I(n), L_I(n)
416
Determine the total effective power, P_eff(n), available for distribution to the eigenmodes in set I(n)
418
i = 1
420
Determine the amount of transmit power, P_i, to allocate to eigenmode i in set I(n)
422
i < L_I(n)?
424
YES
i = i + 1
426
NO
Any P_i < 0?
428
NO
YES
Remove all eigenmodes with P_i < 0 from set I(n)
430
n = n + 1
432
End

FIG. 4



$P_{eff}/3 = P_{set}$
σ^2/λ_{22}^2
σ^2/λ_{33}^2
σ^2/λ_{11}^2
P_1
P_2
P_3
Power
1
2
3
Eigenmodes

FIG. 5A



σ^2/λ_{22}^2
$P_{eff}/3 = P_{set}$
σ^2/λ_{33}^2
σ^2/λ_{11}^2
P_1
P_3
Power
1
2
3
Eigenmodes

FIG. 5B

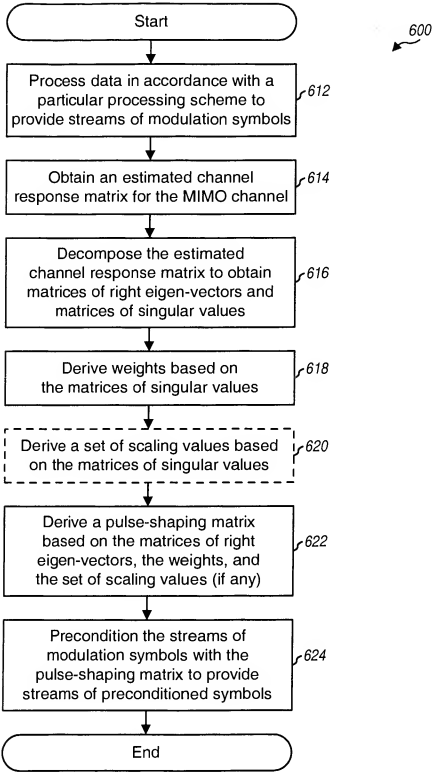

Start
600
Process data in accordance with a particular processing scheme to provide streams of modulation symbols
612
Obtain an estimated channel response matrix for the MIMO channel
614
Decompose the estimated channel response matrix to obtain matrices of right eigen-vectors and matrices of singular values
616
Derive weights based on the matrices of singular values
618
Derive a set of scaling values based on the matrices of singular values
620
Derive a pulse-shaping matrix based on the matrices of right eigen-vectors, the weights, and the set of scaling values (if any)
622
Precondition the streams of modulation symbols with the pulse-shaping matrix to provide streams of preconditioned symbols
624
End

FIG. 6

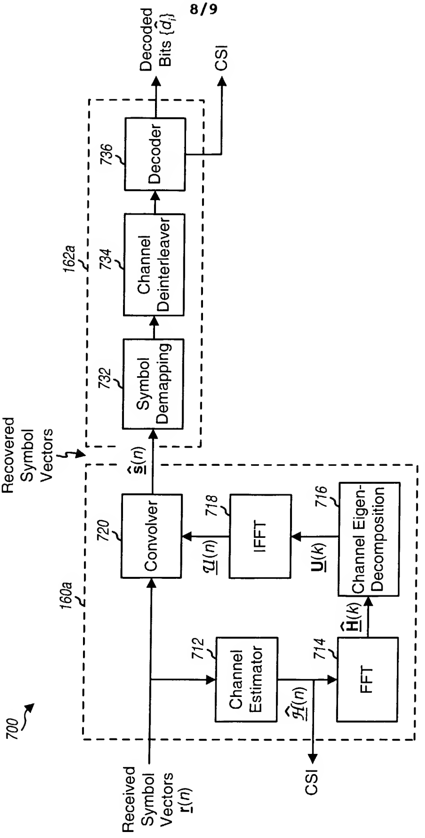

700
Received Symbol Vectors r(n)
160a
Channel Estimator
712
$\hat{\mathcal{H}}(n)$
CSI
FFT
714
$\hat{\mathbf{H}}(k)$
Channel Eigen-Decomposition
716
U(k)
IFFT
718
$\mathcal{U}(n)$
Convolver
720
$\hat{\mathbf{s}}(n)$
Recovered Symbol Vectors
162a
Symbol Demapping
732
Channel Deinterleaver
734
Decoder
736
Decoded Bits $\{\hat{d}_i\}$
CSI

FIG. 7

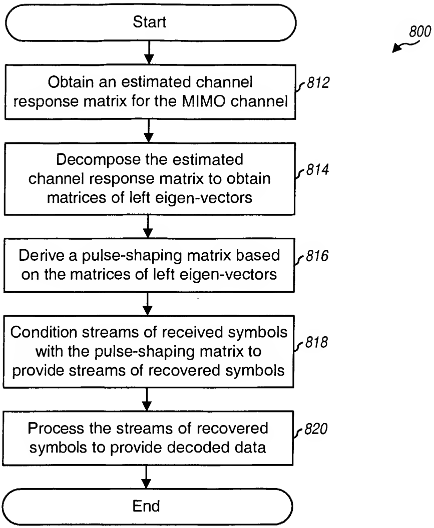

800
Start
Obtain an estimated channel response matrix for the MIMO channel
812
Decompose the estimated channel response matrix to obtain matrices of left eigen-vectors
814
Derive a pulse-shaping matrix based on the matrices of left eigen-vectors
816
Condition streams of received symbols with the pulse-shaping matrix to provide streams of recovered symbols
818
Process the streams of recovered symbols to provide decoded data
820
End

FIG. 8

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 03/19464

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04L1/06 H04L25/03 H04L25/02

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SAMPATH H ET AL: "Joint transmit and receive optimization for high data rate wireless communication using multiple antennas" SIGNALS, SYSTEMS, AND COMPUTERS, 1999. CONFERENCE RECORD OF THE THIRTY-THIRD ASILOMAR CONFERENCE ON OCT. 24-27, 1999, PISCATAWAY, NJ, USA,IEEE, US, 24 October 1999 (1999-10-24), pages 215-219, XP010373976 ISBN: 0-7803-5700-0 abstract page 215, paragraph 2 page 215, paragraph 5 -page 216, paragraph 2 page 217, column 2 -page 218, paragraph 1	1,13-18, 21-30, 35-40
A		2-12,19, 20,31-34
	--- -/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 September 2003	06/10/2003

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Reilly, D

Form PCT/ISA/210 (second sheet) (July 1992)

International Application No
PCT/US 03/19464

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	WO 02 05506 A (QUALCOMM INC) 17 January 2002 (2002-01-17) abstract page 13, line 4 - line 11 page 14, line 24 -page 19, line 4	1,13-18, 21-30, 35-40
A		2-12,19, 20,31-34

X	JOONSUK KIM ET AL: "Transmission Optimization with a Space-Time Filter at Low SNR Wireless Environment" GLOBECOM'99, vol. 1B, 5 December 1999 (1999-12-05), pages 889-893, XP010373673 abstract page 890, column 1, last paragraph -page 892, column 1, paragraph 2	1,13-18, 21-30, 35-40
A		2-12,19, 20,31-34

X	BURR A G: "Adaptive space-time signal processing and coding" IEEE, vol. 2, 22 October 2000 (2000-10-22), pages 710-714, XP010531996 abstract page 710, column 2, paragraph 3 -page 712, column 1, paragraph 3	1,13-18, 21-30, 35-40
A		2-12,19, 20,31-34

A	EP 0 549 019 A (KONINKL PHILIPS ELECTRONICS NV) 30 June 1993 (1993-06-30) abstract page 2, line 25 - line 30	2-12,19, 20,31-34

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 03/19464

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 0205506	A	17-01-2002	AU	7341101 A	21-01-2002
			EP	1299978 A2	09-04-2003
			WO	0205506 A2	17-01-2002
EP 0549019	A	30-06-1993	DE	69227356 D1	26-11-1998
			DE	69227356 T2	27-05-1999
			EP	0549019 A2	30-06-1993
			JP	5327558 A	10-12-1993
			KR	260809 B1	01-07-2000
			US	5353310 A	04-10-1994

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 February 2004 (19.02.2004)

PCT

(10) International Publication Number
WO 2004/016007 A1

(51) International Patent Classification[7]: H04Q 7/20

(21) International Application Number: PCT/US2003/024889

(22) International Filing Date: 7 August 2003 (07.08.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/401,920 8 August 2002 (08.08.2002) US
10/324,194 20 December 2002 (20.12.2002) US

(71) Applicant (*for all designated States except US*): **FLARION TECHNOLOGIES, INC.** [US/US]; Bedminster One, 135 Route 202/206 South, Bedminster, NJ 07921 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **LAROIA, Rajiv** [IN/US]; 455 Somerville Road, Basking Ridge, NJ 07920 (US). **LI, Junyi** [CN/US]; 357 Wren Lane, Bedminster, NJ 07921 (US). **UPPALA, Sathyadev, Venkata** [IN/US]; 281 Spruce Mill Lane, Scotch Plains, NJ 07076 (US).

(74) Agent: **STRAUB, Michael, P.**; Straub & Pokotylo, Suite 83, Building 6, 1 Bethany Road, Hazlet, NJ 07730 (US).

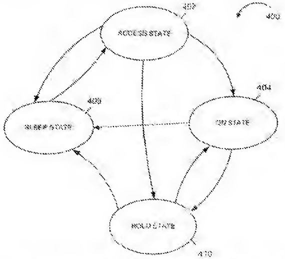
(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,*

[Continued on next page]

(54) Title: METHOD AND APPARATUS FOR OPERATING MOBILE NODES IN MULTIPLE STATES



(57) Abstract: The use of multiple states of mobile communication device (14) operation to allow a single base station (12) to support a relatively large number of mobile nodes (14, 16) is described. The various states require different amounts of communications resources, e.g., bandwidth. Four supported states of operation are a on-state (404), a hold-state (410), a sleep-state (408), and an access-state (402). Each mobile node in the on-state (404) is allocated communication resources to perform transmission power control signaling, transmission timing control signaling and to transmit data as part of a data uplink communications operation. Each mobile node in the hold-state (410) is allocated communication resources to perform transmission timing control signaling and is provided a dedicated uplink for requesting a state transition and a shared resource for transmitting acknowledgements. In the sleep state (408) a mobile node is allocated minimal resources and does not conduct power control signaling or timing control signaling

WO 2004/016007 A1

MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

Published:

— with international search report

— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHODS AND APPARATUS FOR OPERATING MOBILE NODES IN MULTIPLE STATES

RELATED APPLICATION

This application claims the benefit of U.S. Patent Application S.N. 10/324,194 filed December 20, 2002 titled "Methods and Apparatus for Operating Mobile Nodes in Multiple States" and the benefit of U.S. Provisional Patent Application S.N. 60/401,920 filed on August 8, 2002, titled: "Methods and Apparatus for Implementing Mobile Communications System" which is hereby expressly incorporated by reference.

FIELD OF THE INVENTION

The present invention is directed to wireless communications systems and, more particularly, to methods and apparatus for supporting a plurality of mobile nodes in a communications cell with limited resources.

BACKGROUND OF THE INVENTION

Wireless communications systems are frequently implemented as one or more communications cells. Each cell normally includes a base station which supports communications with mobile nodes that are located in, or enter, the communications range of the cell's base station. Within a cell or a sector of a cell, the unit of communications resource is a symbol, e.g., QPSK or QAM transmitted on one frequency tone for one time slot in an orthogonal frequency division multiplexed (OFDM) system. The total available communication resource is divided into a number of such symbols (units) which can be used for communicating control and data information between a base station and one or more mobile nodes in the cell and tends to be limited. Control signals transmitted between a basestation and a mobile node may be transmitted in two possible directions, i.e., from the basestation to the mobile node or from the mobile node to the base station. Transmission of signals from the base station to the mobile is often called a downlink. In contrast, transmission from the mobile to the base station is commonly referred to as an uplink.

In order to provide efficient use of limited communications resources, base stations may allocate different numbers of tones to different mobile nodes depending on the devices' bandwidth needs. In a multiple access system, several nodes may be transmitting data, e.g., in the form of symbols, to a base station at the same time using different tones. This is common in OFDM systems. In such systems, it is important that symbols from different mobile nodes arrive at the base station in a synchronized manner, e.g., so the base station can properly determine the symbol period to which a received symbol belongs and signals from different mobile nodes do not interfere with each other. As mobile nodes move in a cell, transmission delay will vary as a function of the distance between a mobile node and a base station. In order to make sure that transmitted symbols will arrive at a base station from different mobile nodes in synchronized manner, timing control signals, e.g., feedback signals, may be and in many cases are, transmitted to each active mobile node of a cellular system. The timing control signals are often specific to each device and represent, e.g., timing corrections of offsets to be used by the device to determine symbol transmission timing. Timing control signaling operations include, e.g., monitoring for timing control signals, decoding received timing control signals, and performing timing control update operations in response to the decoded received timing control signals.

Timing control signals can be particularly important in systems where there are a large number of mobile nodes. In order to avoid interference from a mobile node due to timing miss synchronization, it may be necessary to establish timing synchronization and control before allowing a mobile node to transmit data, e.g., voice data, IP packets including data, etc. to a base station.

In addition to managing limited resources such as bandwidth, power management is often a concern in wireless communications systems. Mobile nodes, e.g., wireless terminals, are often powered by batteries. Since battery power is limited, it is desirable to reduce power requirements and thereby increase the amount of time a mobile node can operate without a battery recharge or battery replacement. In order to minimize power consumption, it is desirable to limit the amount of power used to transmit signals to a base station to the minimal amount of power required. Another advantage of minimizing mobile node transmission power is that it has the additional benefit of limiting the amount of interference that the transmissions will cause in neighboring cells which will often use the same frequencies as an adjoining cell.

In order to facilitate transmission power control, power control signaling, e.g., a feedback loop, may be established between a base station and a mobile node. Power control signaling often takes place at a much faster rate than the timing control signaling. This is because power control signaling attempts to track variations in the signal strength between the base station and the mobile nodes and this can typically vary on the scale of milliseconds. The timing control needs to take into consideration changes in the distance between base station and the mobile nodes and this tends to vary on a much slower scale, typically hundreds of milliseconds to seconds. Thus the amount of control signaling overhead for power control tends to be much more than that for timing control.

In addition to timing and power control signaling, other types of signaling may be employed. For example mobile nodes in addition may also signal on an uplink the quality of the downlink channel. This may be used by the base station to determine the communication resource allocation to allow for the transfer of data packets from the base station to the mobile. Such downlink channel quality reports allows a base station to determine which mobile node to transmit to and if a mobile node is chosen then the amount of forward error correction protection to apply to the data. These downlink channel quality reports generally are signaled on a similar time scale as the power control signaling. As another example, signaling may be used to periodically announce a mobile node's presence in a cell to a base station. It can also be used to request allocation of uplink resources, e.g., to transmit user data in a communications session. Shared as opposed to dedicated resources may be used for such announcements and/or resource requests.

Signaling resources, e.g., time slots or tones, may be shared or dedicated. In the case of shared time slots or tones, multiple devices may attempt to use the resource, e.g., segment or time slot, to communicate information at the same time. In the case of a shared resource, each ode in the system normally tries to use the resource on an as needed basis. This sometimes results in collisions. In the case of dedicated resources, e.g., with time slots and/or tones being allocated to particular communications device or group of devices to the exclusion of other devices for a certain period of time, the problem of possible collisions is avoided or reduced. The dedicated resources may be part of a common resource, e.g., a common channel, where segments of the channel are dedicated, e.g., allocated, to individual devices or groups of devices where the groups include fewer than the total number of mobile nodes in a cell. For example, in the case of an uplink time segments may be dedicated to individual mobile nodes to

prevent the possibility of collisions. In the case of a downlink, time slots may be dedicated to individual devices or, in the case of multicast messages or control signals, to a group of devices which are to receive the same messages and/or control signals. While segments of a common channel may be dedicated to individual nodes at different times, over time multiple nodes will use different segments of the channel thereby making the overall channel common to multiple nodes.

A logical control channel dedicated to an individual mobile node may be comprised of segments of a common channel dedicated for use by the individual mobile node.

Dedicated resources that go unused may be wasteful. However, shared uplink resources which may be accessed by multiple users simultaneously may suffer from a large number of collisions leading to wasted bandwidth and resulting in an unpredictable amount of time required to communicate.

While timing and power control signals and downlink channel quality reports are useful in managing communications in a wireless communications system, due to limited resources it may not be possible for a base station to support a large number of nodes when power control, and other types of signaling need to be supported on a continuous basis for each node in the system.

In view of the above discussion, it is apparent that there is a need for improved methods of allocating limited resources to mobile nodes to permit a relatively large number of nodes to be supported by a single base station with limited communications resources. It is desirable that at least some methods of communications resource allocation and mobile node management take into consideration the need for timing control signaling and the desirability of power control signaling in mobile communications systems.

SUMMARY OF THE INVENTION

The present invention is directed to methods and apparatus for supporting multiple wireless terminals, e.g., mobile nodes, using a single base station and limited resources such as bandwidth for the transmission of signals between the base station and mobile nodes, e.g., in a communications cell. A system may be implemented in accordance with the invention as a plurality of cells, each cell including at least one base station which serves multiple mobile nodes. A mobile node can, but need not, move within a cell or between cells.

In accordance with the present invention, mobile nodes support multiple states of operation. The control signaling resources used by a mobile node vary depending on the state of operation. Thus, depending on the state of the mobile node, a large amount of signaling resources may be required while in other states a minimum amount of resources may be required. Control signaling resources are in addition to data transmission resources, e.g., bandwidth used to communicate payload data such as voice, data files, etc. By supporting different mobile node states of operation, requiring differing amounts of base station/mobile node control communications resources, e.g., signal bandwidth, used for control purposes, more mobile nodes can be supported by a base station than could be supported if all mobile nodes were allocated the same amount of communications resources for control signaling purposes.

Bandwidth allocated to a particular mobile device for communicating control signals between the mobile device and a base station is known as dedicated control bandwidth. Dedicated control bandwidth may comprise multiple dedicated logical or physical control channels. In some embodiments, each dedicated control channel corresponds to one or more dedicated segments of a common control channel. Control channel segments may be, e.g., channel time slots used for transmitting and/or receiving control signals. Dedicated uplink control channel segments differ from shared uplink control channel segments where multiple devices share the same bandwidth for uplink signaling.

In the case of a shared communications channel conflicts may result when multiple nodes, at the same time attempt to transmit a control signal using the shared communications channel.

Mobile nodes implemented in accordance with one exemplary embodiment support four states, e.g. modes of operation. The four states are a sleep state, a hold state, an access state, and an on state. Of these the access state is a transitory stage and the other states are steady states and the mobile nodes can be in these states for an extended period of time.

Of the four states, the on state requires the highest amount of control signaling resources, e.g., bandwidth used for control signaling purposes. In this state, the mobile node is allocated bandwidth on as needed basis for transmitting and receiving traffic data, e.g., payload information such as text or video. Thus, at any given time in the on state a mobile node may be allocated a dedicated data channel for transmitting payload information. In the on state the mobile node is also allocated a dedicated uplink control signaling channel.

In various embodiments, a dedicated uplink control channel is used during the on state by the MN to make downlink channel quality reports, communicate resource requests, implement session signaling, etc. Downlink channel quality reports are normally signaled frequently enough to track variations in the signal strength between the base station and the mobile node.

During the on state, the base station and mobile node exchange timing control signals using one or more dedicated control channels allowing the mobile node to periodically adjust its transmission timing, e.g., symbol timing, to take into consideration changes in distance and other factors which might cause the transmitted signals to drift timing from the base station's perspective, with the signals transmitted by other mobile nodes. As discussed above, the use of timing control signaling and performing timing control signaling operations, such as updating transmission timing, is important in many systems which use orthogonal frequency division multiple access in the uplink to avoid interference from transmission signals generated by multiple nodes in the same cell.

To provide transmission power control, during the on state, transmission power control signaling is employed to provide a feedback mechanism whereby a mobile node is able to efficiently control its transmission power levels based on signals periodically received from the base station with which it is communicating. In various embodiments the base station periodically transmits power control signals over a dedicated control downlink. As part of the transmission power control signaling process, the mobile node, performs various transmission

power control signaling operations including, for example, monitoring for transmission power control signals directed to the particular mobile node, decoding received transmission power control signals, and updating its transmission power levels based on the received and decoded transmission power control signals. Thus, in response to receiving power control signals in a dedicated downlink segment corresponding to the particular mobile node, the mobile node adjusts its transmission power level in response to the received signal. In this manner, a mobile node can increase or decease its transmission power to provide for successful receipt of signals by the base station without excessive wastage of power and therefore reducing interference and improving battery life. The power control signaling is typically carried out sufficiently frequently to track fast variations in the signal strength between the base station and the mobile nodes. The power control interval is a function of smallest channel coherence time that the system is designed for. The power control signaling and the downlink channel quality reports are normally of similar time scale, and in general, occur at a much higher frequency than the timing control signaling. However, in accordance with one feature of the present invention the base station varies the rate at which it transmits power control signals to a mobile node as a function of the mobile node's state of operation. As a result, in such an embodiment, the rate at which the mobile node performs transmission power control adjustments will vary as a function of the state in which the mobile node operates. In one exemplary embodiment, power control updates are not performed in the sleep state and, when performed in the hold state, are normally performed at a lower rate than during the on state.

Operation of a mobile node in the hold state requires fewer control communications resources, e.g., bandwidth, than are required to support operation of a mobile node in the on state. In addition, in various embodiments while in the hold state a mobile node is denied bandwidth for transmitting payload data, but the mobile can be allocated bandwidth for receiving payload data. In such embodiments the mobile node is denied a dedicated data uplink communications channel during the hold state. The bandwidth allocated for receiving data may be, e.g., a data downlink channel shared with other mobile nodes. During the hold state timing control signaling is maintained and the mobile node is also allocated a dedicated control uplink communication resource, e.g., dedicated uplink control communications channel, which it can use to request changes to other states. This allows, for example, a mobile node to obtain additional communications resources by requesting a transition to the on state where it could transmit payload data. In some but not all embodiments, in the hold state, the dedicated uplink control channel is limited to the communication of signals requesting permission to change the

state of mobile node operation, e.g., from the hold state to the on state. During the hold state the bandwidth allocated, e.g., dedicated, to a mobile node for control signaling purposes is less than in the on-state.

Maintaining timing control while in the hold-state allows the mobile nodes to transmit their uplink requests without generating interference to other mobiles within the same cell and having a dedicated uplink control resource ensures that the delays for state transition are minimal as the requests for state transitions do not collide with similar requests from other mobile nodes as may occur in the case of shared uplink resources. Since timing control signaling is maintained, when the mobile node transitions from the hold state to the on state it can transmit data without much delay, e.g., as soon as the requested uplink resource is granted, without concerns about creating interference for other mobile nodes in the cell due to drift of uplink symbol timing. During the hold state, transmission power control signaling may be discontinued or performed less frequently, e.g., at greater intervals than performed during on state operation. In this manner, the dedicated control resources used for power control signaling can be eliminated or reduced allowing fewer resources to be dedicated to this purpose than would be possible if power control signaling for all nodes in the hold state was performed at the same rate as in the on state.

When transitioning from the hold state to the on state, the mobile node may start off with an initial high power level to insure that its signals are received by the base station with the power level being reduced once transmission power control signaling resumes at a normal rate as part of on state operation. In one exemplary embodiment, when the mobile node in the hold state intends to migrate to the on state, it transmits a state transition request using a dedicated uplink communication resource, which is not shared with any other mobile nodes. The base station then responds with a broadcast message indicating its response to the mobile's state transition request. The mobile on receiving the base station message meant for it responds with an acknowledgement. The acknowledgment is transmitted over a shared resource on the uplink and is slaved to the broadcast message on the downlink.

By transmitting an appropriate state transition request the mobile may also transition to the sleep state. In one exemplary embodiment, when the mobile node does not intend to migrate to another state, the mobile node may not transmit any signal in its dedicated uplink communication channel, though the dedicated channel has been assigned to the mobile

node and is therefore not used by any other mobile nodes. In another embodiment, the mobile node uses an on/off signaling in its dedicated uplink communication channel, where the mobile node sends a fixed signal (on) when it intends to migrate to another state and does not send any signal (off) when it does not intend to migrate to any other state. In this case, the transmission of the fixed signal can be interpreted as a migration request to the on state if the transmission occurs at certain time instances, and as a migration request to the sleep state if the transmission occurs at some other time instances.

In order to support a large number of mobile nodes, a sleep state requiring relatively few communications resources is also supported. In an exemplary embodiment, during the sleep state, timing control signal and power control signaling are not supported. Thus, in the sleep state, the mobile nodes normally do not performing transmission timing control or transmission power control signaling operations such as receiving, decoding and using timing and transmission power control signals. In addition, the mobile node is not allocated a dedicated uplink control resource, e.g., uplink control communications channel, for making state transition requests or payload transmission requests. In addition, during the sleep state the mobile node is not allocated data transmission resources, e.g., dedicated bandwidth, for use in transmitting payload data, e.g., as part of a communications session with another node conducted through the base station.

Given the absence of a dedicated uplink control channel during the sleep state, a shared communications channel is used to contact the base station to request resources necessary for a mobile node to initiate transition from the sleep state to another state.

In some embodiments, in the sleep state the mobile node may, at the behest of the base station serving the cell, signal its presence in the cell, e.g., using a shared communications resource. However, as discussed above, little other signaling is supported during this state of operation. Thus, very little control signaling bandwidth is used to communicate control information between mobile nodes in the sleep state and a base station serving the nodes.

The access state is a state through which a node in the sleep state can transition into one of the other supported states. The transition between states may be triggered by an action by a user of the mobile node, e.g., an attempt to transmit data to another mobile node. Upon entering the access state, transmission power control and timing control signaling has not

yet been established. During access state operation, timing control signaling is established and, in various embodiments, full or partial transmission power control signaling is established. A mobile node can transition from the access state to either the on state or the hold state.

The establishment of the timing synchronization and transmission power control can take some amount of time during which data transition is delayed. Also the access process happens through a shared media and contentions between mobile nodes need to be resolved. By supporting a hold state in accordance with the present invention, in addition to a sleep state, such delays can be avoided for a number of mobile nodes, as transition from the hold state to the on state does not go through the access state, while the number of nodes which can be supported by a single base station is larger than would be possible without the use of reduced signaling states of mobile node operation.

In some embodiments, for an individual cell, the maximum number of mobile nodes that can be in the sleep state at any given time is set to be greater than the maximum number of mobile nodes that can be in the hold state at given time. In addition, the maximum number of mobile nodes which can be in the hold at any given time is set to be greater than the maximum number of nodes that can be in the on state at any given time.

In accordance with a power conservation feature of the present invention, downlink control signaling from the base station to the mobile nodes is divided into a plurality of control channels. A different number of downlink control channels are monitored by a mobile node depending on the node's state of operation. During the on state the greatest number of downlink control channels are monitored. During the hold state a smaller number of downlink control channels are monitored than during the on state. In the sleep state the smallest number of downlink control channels are monitored.

To further reduce power consumption in the mobile node associated with monitoring for control signals, in accordance with one feature of the invention control channels monitored during the hold and sleep states are implemented as periodic control channels. That is, signals are not broadcast on a continuous basis on the control channels monitored in the hold and sleep states. Thus, during the hold and sleep states the mobiles monitor for control signals at periodic intervals and save power by not monitoring for control signals at those times when control signals are not transmitted on the monitored channels. To further decrease the time a

particular mobile needs to monitor for control signals during the hold and sleep states, portions, e.g., segments, of the periodic control channels may be dedicated to one or a group of mobile nodes. The mobile nodes are made aware of which control channel segments are dedicated to them and then monitor the dedicated segments as opposed to all the segments in the control channels. This allows monitoring for control signals to be performed in the hold and sleep states by individual mobile nodes at greater periodic intervals than would be possible if the mobile were required to monitor all segments of the periodic control channels.

In one particular embodiment, during the on state, mobile nodes monitor segments of an assignment channel on a continuous basis and also monitor segments of periodic fast paging and slow paging control channels. When in the hold state the mobiles monitor the fast paging and slow paging control channels. Such monitoring may involve monitoring a subset of the segments of the periodic fast and slow paging channels, e.g., segments dedicated to the particular mobile node. During the hold state in the particular exemplary embodiment the slow paging channel is monitored but not the fast paging channel or the assignment channel. The paging control channels may be used to instruct the mobile node to change states.

By limiting the number of control channels and the rate of control channel monitoring as a function of the state of operation, power resources can be conserved in accordance with the invention while operating in the hold and sleep states.

Numerous additional features, benefits and details of the methods and apparatus of the present invention are described in the detailed description which follows.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 illustrates an exemplary communication cell, which may be part of a communications system, implemented in accordance with the present invention.

Fig. 2 illustrates a base station implemented in accordance with the present invention.

Fig. 3 illustrates a mobile node implemented in accordance with the present invention.

Fig. 4 is a state diagram illustrating the different states that a mobile node may enter while operating in accordance with the present invention.

Fig. 5 is a chart illustrating various control and signaling modules that are executed by a mobile node during each of the different states illustrated in Fig. 4.

Fig. 6 illustrates the transmissions associated with three exemplary downlink control channels used in accordance with one embodiment of the present invention.

Fig. 7 illustrates which control channels shown in Fig. 6 are monitored in each of the four states in which a mobile node of the present invention may operate.

DETAILED DESCRIPTION

Fig. 1 illustrates a communications cell 10 implemented in accordance with the present invention. A communications system may include multiple cells of the type illustrated in Fig. 1. The communications cell 10 includes a base station 12 and a plurality, e.g., a number N, of mobile nodes 14, 16 which exchange data and signals with the base station 12 over the air as represented by arrows 13, 15. In accordance with the invention, the base station 12 and mobile nodes 14, 16 are capable of performing and/or maintaining control signaling independently of data signaling, e.g., voice or other payload information, being communicated. Examples of control signaling include power control, downlink channel quality reports, and timing control signaling.

Fig. 2 illustrates a base station implemented in accordance with the present invention. As shown, the exemplary BS 12 includes a receiver circuit 202, transmitter circuit 204, processor 206, memory 210 and a network interface 208 coupled together by a bus 207. The receiver circuit 202 is coupled to an antenna 203 for receiving signals from mobile nodes. The transmitter circuit 204 is coupled to a transmitter antenna 205 which can be used to broadcast signals to mobile nodes. The network interface 208 is used to couple the base station 12 to one or more network elements, e.g., routers and/or the Internet. In this manner, the base station 12 can serve as a communications element between mobile nodes serviced by the base station 12 and other network elements.

Operation of the base station 12 is controlled by the processor 206 under direction of one or more routines stored in the memory 210. Memory 210 includes communications routines 223, data 220, session management/resource allocation routine 222, session and resource signaling subroutine 224, and active user information 212. Communications routines 223, include various communications applications which may be used to provide particular services, e.g., IP telephony services or interactive gaming, to one or more mobile node users. Data 220 includes data to be transmitted to, or received from, one or more mobile nodes. Data 220 may include, e.g., voice data, E-mail messages, video images, game data, etc.

The session management and resource allocation routine 222 operates in conjunction with subroutines 224 and active user information 212 and data 220. The routine 222 is responsible for determining whether and when mobile nodes may transition between states and also the resources allocated to a mobile node within a state. It may base its decision on various criteria such as, requests from mobile nodes requesting to transition between states, idletime/time spent by a mobile in a particular state, available resources, available data, mobile priorities etc. These criteria would allow a base station to support different quality of service (QOS) across the mobile nodes connected to it.

The session and resource signaling subroutine 224 is called by session management routine 222 when signaling operations are required. Such signaling is used to indicate the permission to transition between states. It is also used to allocate the resources, e.g., when in a particular state. For example, in the on state a mobile node may be granted resources to transmit or receive data.

Active user information 212 includes information for each active user and/or mobile node serviced by the base station 12. For each mobile node and/or user it includes a set of state information 213, 213'. The state information 213, 213' includes, e.g., whether the mobile node is in an on state, a hold state, a sleep state, or an access state as supported in accordance with the present invention, number and types of data packets currently available for transmission to or from the mobile node, and information on the communication resources used by the mobile node.

Fig. 3 illustrates an exemplary mobile node 14 implemented in accordance with the invention. The mobile node 14 includes a receiver 302, a transmitter 304, antennas 303, 305, a memory 210 and a processor coupled together as shown in Fig. 3. The mobile node uses its transmitter 306, receiver 302, and antennas 303, 305 to send and receive information to and from base station 12.

Memory 210 includes user/device information 312, data 320, a power control and power control signaling module 322, a timing control and timing control signaling module 324, a device status control and status signaling module 326, and a data control and data signaling module 328. The mobile node 14 operates under control of the modules, which are executed by the processor 306. User/device information 312 includes device information, e.g., a device identifier, a network address or a telephone number. This information can be used, by the base station 12, to identify the mobile nodes, e.g., when assigning communications channels. The user/device information 312 also includes information concerning the present state of the mobile device 14. The data 320 includes, e.g., voice, text and/or other data received from, or to be transmitted to, the base station as part of a communications session.

Device status control and status signaling module 326 is used for device status control and status signaling. Device status control module 326 determines, in conjunction with signals received from the base station 12, what mode, e.g., state, the mobile node 14 is to operate in at any given time. In response to, e.g., user input, the mobile node 14 may request permission from the base station 12 to transition from one state to another and to be granted the resources associated with a given state. Depending on the state of operation at any given time and the communications resources allocated to the mobile node 14, status control and status signaling module 326 determines what signaling is to occur and which signaling modules are to be active. In response to periods of reduced signal activity, e.g., control signal activity, status control and status signaling module 326 may decide to transition from a current state of operation to a state of operation requiring fewer control resources and/or requires less power. The module 326 may, but need not, signal the state transition to the base station. Status control and status signaling module 326 controls, among other things, the number of downlink control channels monitored during each state of operation and, in various embodiments, the rate at which one or more downlink control channels are monitored.

As part of the processes of controlling the state of the mobile node 14, and overseeing general signaling between the mobile node 14 and base station 12, the signaling module is responsible for signaling to the base station 12, when the mobile node 14 first enters a cell and/or when the base station 12 requests that the mobile node 14 indicate it presence. The mobile node 14 may use a shared communication resource to signal its presence to the cell's base station 12, while a dedicated communication resource may be used for other communication signals, e.g., uploading and downloading data files as part of a communication session.

Transmission power control and power control signaling module 322 is used to control the generation, processing and reception of transmission power control signals. Module 322 controls the signaling used to implement transmission power control through interaction with the base station 12. Signals transmitted to, or received from the base station 12 are used to control mobile node transmission power levels under direction of module 322. Power control is used by the base station 12 and the mobile nodes 14, 16 to regulate power output when transmitting signals. The base station 12 transmits signals to the mobile nodes which are used by the mobile nodes in adjusting their transmission power output. The optimal level of power used to transmit signals varies with several factors including transmission burst rate, channel conditions and distance from the base station 12, e.g., the closer the mobile node 14 is to the base station 12, the less power the mobile node 14 needs to use to transmit signals to the base station 12. Using a maximum power output for all transmissions has disadvantages, e.g., the mobile node 14 battery life is reduced, and high power output increases the potential of the transmitted signals causing interference, e.g., with transmissions in neighboring or overlapping cells. Transmission power control signaling allows the mobile node to reduce and/or minimize transmission output power and thereby extend battery life.

Timing control and timing control signaling module 324 is used for timing and timing signaling. Timing control is used in wireless networking schemes such as, e.g., those with uplinks based on orthogonal frequency division multiple access. To reduce the effects of noise, tone hopping may also be used. Tone hopping may be a function of time with different mobile nodes being allocated different tones during different symbol transmission time periods, referred to as symbol times. In order for a base station 12 of a multiple access system to keep track of, and distinguish between, signals from different mobile nodes, it is desirable for the base station 12 to receive information from the mobile nodes in a synchronized manner. A drift of

timing between the mobile node 14 and the base station 12 can cause transmission interference making it difficult for the base station to distinguish between symbols transmitted by different mobile nodes, e.g., using the same tone, but during different symbol time periods or using different tones but during the same symbol time period.

For example, the effect on a mobile node's distance from the base station is a factor since transmissions from mobile node that are farther from the base station 12 take longer to reach the base station 12. A late arriving signal can interfere with another connection that has hopped to the late arriving signal's frequency in a latter time period. In order to maintain symbol timing synchronization, it is required to instruct a node to advance or delay its symbol transmission start time to take into consideration changes in signal propagation time to the base station.

Data and data signaling module 328 is used to control transmission and the reception of payload data, e.g., a channel or time slot dedicated to the mobile node for signaling purposes. This includes, e.g., the data packets of an Internet file transfer operation.

In accordance with the present invention, the mobile node 14 can be in one of four states. The signaling, power, and communications resources required by a mobile node will vary depending on the sate in which the mobile node is operating. As a result of using multiple states in the mobile nodes, the base station 12 is able to allocate different degrees of communication resource, e.g., control and data signaling resource, to different mobile nodes as a function of the node's state of operation. This allows the base station 12 to support a greater number of mobile nodes than would be possible if all nodes were continuously in the on state. The particular state that the mobile node 14 is in determines the control signaling and data signaling modules that are executed at any given time and also the level of control signaling between the mobile node and base station 12. The mobile node 14 can also take advantage of the different activity level in different states to save power and extend battery life.

Operation of the mobile nodes 14 in different states, in accordance with the present invention, will now be explained with reference to figures 4 and 5. Fig. 4 illustrates a state diagram 400 including four possible states, an access state 402, a on state 404, a hold state 410 and a sleep state 408, that a mobile node 14 can enter. Arrows are used in Fig. 4 to illustrate the possible transitions between the four states.

Fig. 5 illustrates the mobile node modules 322, 324, 326, 328 that are in the various states shown in Fig. 4. Each row of the chart 500 corresponds to a different state. The first through fourth rows 502, 504, 506, 508 correspond to the sleep state, access state, on state, and hold state, respectively. Each column of the chart 500 corresponds to a different module within the mobile node 14. For example, the first column 510 corresponds to the transmission power control and power control signaling module 322, the second column 512 corresponds to the timing control and timing control signaling module 324, the third column 514 corresponds to the device status control and status signaling module 326, while the last column 516 corresponds to the data and data signaling module 328. In Fig. 5, solid lines are used to indicate modules which are active in a particular state. Short dashed lines are used to indicate modules which may transition from an inactive or reduced activity level to a fully active status before the access state is exited, assuming the modules are not already fully active. Long dashed lines are used to indicate a module which may be active in a state but which may perform signaling at a reduced rate while in the indicated state as opposed to the signaling rate implemented in the on state.

From Fig. 5 it can be seen that during the sleep state the device status control and status signaling module 326 remains active but the other modules are inactive allowing for power conservation and a significantly restricting mobile node activity. In the access state 402, which serves as transition state, transmission power control and power control signaling module 322, timing control and timing control signaling module 324 will become fully active (or active at a reduced rate in the case of the transmission power control and power control signaling module 322 in some embodiments) prior to leaving the access state 402 to enter the on-state 404 or hold state 410. In the on-state, all signaling modules 322, 324, 326, 328 are fully active requiring the most power from the mobile node's perspective and the highest allocation of communication resources, e.g., bandwidth, from the base station's perspective. In the hold state, transmission power control and power control signaling module 322 may be inactive or active at a much reduced signaling rate. Timing control and timing control signaling module 324 remains alive as does the device status control and status signaling module 326. The data and data signaling module 326 is either inactive or operates to implement reduced functionality, e.g., receive data but not transmit data as part of a communication session between various nodes. In this manner, the hold state allows bandwidth and other communications resources to be conserved while, in some cases, allowing the mobile node to receive, e.g., multi-cast signals and/or messages.

Each of the states, and potential transition between states, will now be described in detail with reference to the state diagram of Fig. 4.

Of the four states 402, 404, 410, 408, the on state 404 allows the mobile node to perform the widest range of supported communications activities but requires the highest amount of signaling resources, e.g., bandwidth. In this state 404, which may be thought of as a "fully-on" state, the mobile node 14 is allocated bandwidth on an as needed basis for transmitting and receiving data, e.g., payload information such as text or video. The mobile node 14 is also allocated a dedicated uplink signaling channel which it can use to make downlink channel quality reports, communication resource requests, implement session signaling, etc. To be useful, these downlink channel quality reports should be signaled sufficiently frequently to track variations in the signal strengths received by the mobile nodes.

During the on state 404, under control of module 324, the base station 12 and mobile node 14 exchange timing control signals. This allows the mobile node 14 to periodically adjust its transmission timing, e.g., symbol timing, to take into consideration changes in distance and other factors which might cause the mobile node transmitted signals to drift timing at the base station's receiver, with respect to the signals transmitted by other mobile nodes 16. As discussed above, the use of symbol timing control signaling is employed in many systems which use orthogonal frequency division multiple access in the uplink, to avoid interference from transmission signals generated by multiple nodes in the same cell 10.

To provide transmission power control, during the on state 404, transmission power control signaling is employed, under direction of module 322, to provide a feedback mechanism whereby a mobile node is able to efficiently control its transmission power levels based on signals periodically received from the base station with which it is communicating. In this manner, a mobile node 14 can increase and/or decrease its transmission power to provide for successful receipt of signals by the base station 12 without excessive wastage of power and therefore reduced battery life. The power control signaling is carried out sufficiently frequently to track variations in the signal strength between the base station 12 and the mobile nodes 14, 16 for a certain minimum channel coherence time. The power control interval is a function of channel coherence time. The power control signaling and the downlink channel quality reports

are of similar time scale, and in general, occur at much higher rate than the timing control signaling required to support vehicular mobility.

From the on state 404, the mobile node 14 can transition into either the sleep state 408 or the hold state 410. Each of these states requires reduced communication resources, e.g., bandwidth, to support than does the on state 404. The transition may be in response to user input, e.g., a user terminating a communications session or in response to the loss of communications resources, e.g., bandwidth required to support the transmission and/or receipt of information to be communicated such as voice or data information.

In accordance with the present invention, in the hold state, a mobile node is denied bandwidth for transmitting payload data. However, timing control signaling is maintained and the mobile node is also allocated a dedicated uplink communication resource which it can use to request changes to other states. This allows for instance a mobile node to obtain additional communications resources by requesting a transition to on state where it could transmit payload data. Maintaining timing control during the hold state 410 allows the mobile node 14 to transmit its uplink requests without generating interference to other mobiles 16 within the same cell 10. Having a dedicated resource for transmitting requests to the base station 12 also helps ensure that the delays for state transition are minimal as these requests do not collide with similar requests from other mobiles.

From the hold state 410, the mobile node may transition into the on state 404, e.g., upon being granted a requested communication resource. Alternatively, the mobile node can transition into the sleep state 408. Since timing control signaling is maintained in the hold state 410, when the mobile node transitions to the on state it can transmit data without much delay, e.g., as soon as the requested bandwidth is granted, without concerns about creating interference to the uplink transmission of other mobile nodes in the cell which could result from a timing drift of the mobile node.

During the hold state 410, transmission power control signaling may be discontinued or performed at greater intervals, e.g., at a similar rate as timing control. In this manner, the resource, e.g., base station to mobile node control resource, used for transmission power control signaling can be eliminated or less resource can be dedicated to this purpose than would be possible if power control signaling for all nodes 14, 16 in the hold state was performed

at the same rate as in the on state. The mobile nodes 14, 16 transmission power control updates are performed in the mobile node during the hold state at a reduced rate or not at all, in a manner which corresponds to the reduced transmission power control signaling. When transitioning from the hold state 410 to the on state 404, the mobile node 14 may start off with an initial high power level to insure that its signals are received by the base station 12. The power level is then reduced once transmission power control signaling resumes at a normal (full) rate as part of on state operation.

Transition from hold state can be initiated by base station or by the mobile nodes. The base station may initiate a transition by sending a page over a paging channel meant for the hold state users. In one embodiment, the mobile decodes the paging channel with some prearranged periodicity, to check for base station messages. On finding a page message meant for it, it responds with an acknowledgement. In various embodiments the acknowledgement is transmitted over a shared resource on the uplink and is slaved to the page or grant message on the downlink. The mobile node 14 responds to a state change message by moving to the assigned state specified in the received state change message.

In one embodiment, when the mobile node 14 intends to migrate from the hold state 410 to the on state 404, it transmits a state transition request using its dedicated uplink communications channel, which is not shared with any other mobile nodes 16. Since the channel is not shared, the base station 12 is able to receive the request without interference and promptly grant the request assuming the required resources are available taking into account the priority of the user and/or the applications that the user may be using. The mobile on receiving a grant message meant for it, responds with an acknowledgement. The acknowledgment is transmitted over a shared resource on the uplink and is slaved to the grant message on the downlink.

In one exemplary embodiment, when the mobile node does not intend to migrate to another state from the hold state, the mobile node may not transmit any signal in its dedicated uplink communication resource, though the dedicated resource has been assigned to the mobile node and therefore will not be used by any other mobile nodes. In this case, the mobile node can temporarily shut down the transmission module and related functions thereby conserving power.

In another embodiment, the mobile node uses an on/off signaling in its dedicated uplink communication resource, where the mobile node sends a fixed signal (on) when it intends to migrate to another state or does not send any signal (off) when it does not intend to migrate to any other state. In this case, the transmission of the fixed signal can be interpreted as a migration request to the on state if the transmission occurs at certain time instances and as a migration request to the sleep state if the transmission occurs at some other time instances.

In order to provide reachability for a large number of mobile nodes 14, 16, the sleep state 408, requiring relatively few communications resources, is also supported. The mobile node 14 can transition into the sleep state 408, e.g., in response to user input, a period of inactivity, or a signal from the base station 12, from any of the other supported states 404, 404, 410.

In the sleep state 408 the mobile node 14 may, at the behest of the base station 12, serving the cell 10 signal its presence in the cell 10. However, little other signaling is supported during this state 408 of operation. In the exemplary embodiment, during the sleep state 408, timing control signaling and power control signaling are not supported. In addition, the mobile node is not allocated a dedicated uplink for making resource requests and is not allocated bandwidth for use in transmitting payload data, e.g., as part of a communications session with another node 16 conducted through the base station 12.

Transitions from the sleep state 408 to another state 404, 410 occur by passing through access state 402. A shared (contention based), as opposed to a dedicated uplink, communications channel is used to contact the base station 12 to request resources necessary to transition from the sleep state 408 to another state 402, 404, 410. These transitions could be initiated by the base station on the paging channel or by the mobile nodes 14, 16. Since the communications channel used to request resources to transition from the sleep state is shared, a mobile node 14 may encounter delays before being able to successfully transmit the resource request to the base station 12. This is due to possible collisions with similar requests from other mobile nodes. Such delays are not encountered in regard to transitions from the hold state 410 to the on state due to the use of a dedicated uplink resource for requests while in the hold state 410.

The access state 402 is a state through which a node 14 in the sleep state 408 can transition into one of the other supported states 404, 410. The transition out of the sleep state is normally triggered, by an action by a user of the mobile node 14, e.g., an attempt to transmit data to another mobile node 16 or by the base station 12. Upon entering the access state 402, transmission power control and timing control signaling has not yet been established. During access state operation, timing control signaling is established and, in various embodiments, full or partial transmission power control signaling is established with mobile node transmission output power levels being adjusted accordingly. A mobile node can transition from the access state 402, back to the sleep state 408 or to either the on state 404 or the hold state 410. Transition to the sleep state 408 may occur, e.g., in response to a user canceling a transmission request or a base station 12 denying the node the resources required to complete the transition to the hold or on states 404, 410. Transition from the access state to the on state 404 or hold state 410 normally occurs once the mobile node 14 has restored power and timing synchronization signaling with the base station 12 and has been granted the communications resource or resources required to maintain the state into which the mobile node 14 is transitioning.

The establishment of the timing synchronization and transmission power control signaling, in the access state 402, can take some amount of time during which data transmission is delayed. Furthermore, as noted above, delays may result form the use of a shared resources to request the transition which can produce contentions between mobile nodes which take time to resolve. In addition, because of the use of shared resources in requesting a state transition, it is difficult to prioritize between different nodes requesting state transition.

In some embodiments, for an individual cell 10, the maximum number of mobile nodes 14, 16 that can be in the sleep state 408 at any given time is set to be greater than the maximum number of mobile nodes 14, 16 that can be in the hold state 410 at given time. In addition, the maximum number of mobile nodes 14, 16 which can be in the hold state 410 at any given time is set to be greater than the maximum number of nodes that can be in the on state 404 at any given time.

By supporting a hold state in accordance with the present invention, in addition to a sleep state, such delays can be avoided for a number of mobile nodes 14, 16, as transition from the hold state 410 to the on state 404 does not go through the access state, while the number of

nodes which can be supported by a single base station 12 is larger than would be possible without the use of the reduced signaling hold state.

From a power standpoint it is desirable that the amount of time and thus power a mobile node spends monitoring for control signals be minimized. In order to minimize the amount of time and power a mobile node spends monitoring for control signals, at least some downlink control signaling, i.e., signaling from the base station to one or more mobile nodes, is performed using multiple control channels. In one embodiment of the invention, particularly well suited for use with mobile nodes capable of supporting multiple states of operation, a plurality of control channels are provided for communicating control signals from the base station to the mobile nodes. Each of the plurality of common control channels is divided into a number of segments, e.g., time slots, where each segment is dedicated, e.g., assigned, for use by one or a group of mobile nodes. In this case, a group of mobile nodes may be, e.g., a subset of the mobile nodes in the system which correspond to a multicast message group. In such an embodiment, the control channels are common to multiple nodes, but each segment of a channel is dedicated, e.g., corresponds to, a particular one of the mobile nodes or group of mobile nodes with other mobile nodes being excluded from using the dedicated segments. The dedicated segments of a common control channel corresponding to an individual mobile node represent a dedicated control channel allocated to the individual mobile node.

The pattern of control channel segment allocation is made known to the individual mobile nodes 14, 16 in a cell, e.g., based on information transmitted to each particular node 14, 16 from the base station 12.

To provide particularly efficient control channel signaling, base station to mobile node control signaling may be performed at several different rates, with a different control channel being used for each of the different control channel signaling rates.

In order to minimize the amount of power and resources consumed by the task of monitoring control channels for information relevant to a mobile node, each mobile node need only monitor to detect signals in control channel segments assigned to the particular node. This allows the mobile nodes to schedule control channel monitoring operations so that the control channels need not be monitored on a continuous basis while still allowing the mobile nodes to receive control signals in a timely manner.

In one embodiment which is particularly well suited for use where mobile nodes that support at least an on state, a hold state and a sleep state, three different segmented control channels are used. The three control channels include an assignment control channel, a fast paging control channel, and a slow paging control channel.

The fast paging control channel and slow paging control channel are periodic in nature, e.g., control signals are not transmitted in terms of time on a continuous basis in these channels. Thus, mobile nodes need not spend power and resources monitoring these channels on a continuous basis. In some embodiments, to further reduce the amount of time and power a mobile needs to spend monitoring the fast and slow paging channels, the channels are segmented and the segments are dedicated to particular mobile nodes or groups of mobile nodes.

In order to minimize the amount of power and resources consumed by the task of monitoring control channels for information relevant to a mobile node, each mobile node need only monitor to detect signals in the fast and slow paging control channel segments assigned to the particular node. This allows the mobile nodes to schedule control channel monitoring operations so that the control channels can be monitored on a less frequent basis than would be possible if all segments need to be monitored for control signals.

Fig. 6 illustrates control signals 602, 620, 630 corresponding to exemplary assignment, fast paging and slow paging downlink control channels respectively. The fast paging control channel signal 602 is divided into a plurality of segments, e.g., 1 ms time slots. Transmission in the assignment channel occurs, in the Fig. 6 embodiment, on a continuous basis. For each time slot, there is a corresponding traffic channel segment or segments. Traffic channel segments are allocated by the base station 12 to mobile nodes 14, 16 by transmitting a mobile node identifier or mobile node group identifier in a time slot to indicate that the corresponding traffic segment or segments have been assigned for use to the mobile node(s) corresponding to the transmitted identifier. While in the on state mobile nodes 14, 16 monitor the assignment channel on a continuous basis, e.g., at a rate sufficient to detect the identifier included in each segment of the control channel used for traffic assignment purposes.

During the on state, in addition to the assignment channel each mobile node 14, 16 monitors the periodic fast paging and slow paging channels.

In Fig. 6, fast paging signal 620 can be seen to be periodic in nature. Each exemplary fast paging signal period 622, 626, 230, 634 is 10 ms in duration. However, of this 10 ms period, the fast paging signal is actually transmitted for only a fraction of the full period, e.g., 2 ms. The periods 623, 627, 631, 635 in which the fast paging signal is transmitted are segmented into time slots. The remaining portions 624, 628, 632, 636 represent portions of time in which the fast paging control signal is not broadcast by the base station 12. While only two 1 ms segments are shown in each fast paging on period 623, 627, 631, 635 it is to be understood that there are normally several segments per on period.

To reduce the amount of time mobile nodes 14, 16 need monitor for fast paging control signals, fast paging control channel segments are, in some embodiments, dedicated to individual mobile nodes or groups of mobile nodes. The information on which segments are dedicated to which mobile nodes is normally conveyed to the mobile nodes 14, 16, e.g., form the base station 12. Once the dedication information is known, the mobile nodes 14, 16 can limit their monitoring of fast paging channel segments to segments which are dedicated to them. In such embodiments, mobile nodes can monitor the fast paging channel at periodic intervals greater than the fast paging period without risking missing control information transmitted to the mobile on the fast paging channel.

The segments of the fast paging channel are used to convey information, e.g., commands, used to control the mobile node to transition between states. The segments of the fast paging channel can also be used to instruct the mobile node to monitor the assignment channel, e.g., when the mobile node is in a state which has caused it to stop monitoring the assignment channel. Since the mobile nodes of the system know which segments of the fast paging channel are assigned to them, commands may be included in the fast paging channel segments without mobile node identifiers making for an efficient transmission scheme.

The slow paging channel is segmented and used to convey information in the same manner as the fast paging channel. The information conveyed using the slow paging channel may be the same as, or similar to, the information and commands that are transmitted using the fast paging channel.

In Fig. 6, signal 630 represents an exemplary slow paging channel signal. Note that the full slow paging signal period 632 is longer than the paging period 622 of the fast paging channel. Reference numbers 631 and 634 are used in Fig. 6 to show portions of a slow paging period. Given that the slow paging period is longer than the fast paging period, the time between control signal transmission in the slow paging channel tends to be greater than in the fast paging channel. This means that the mobile node may discontinue monitoring the slow paging channel for longer intervals than is possible with the fast paging channel. It also implies, however, that it may take, on average, longer for a control signal transmitted on the slow paging channel to be received by the intended mobile node.

In Fig. 6, two slow paging signal transmission on signal periods 640, 642 are shown. Signal periods 639, 641, 643 correspond to slow paging channel signal periods during which no slow paging signal is transmitted.

Since the fast and slow paging channels are period in nature, if the transmission on periods are staggered so that they do not overlap, the fast and slow paging channels may be implemented using the same physical transmission resources, e.g., tones, with the tones being interpreted as corresponding to either the fast or slow paging channel depending on the time period to which the tones correspond.

The spacing between segments allocated to a particular mobile node in the slow paging channel are often, but need not be, greater than in the fast paging channel. This generally means, in terms of time, that a mobile device needs to monitor the slow paging channel at intervals which are more widely spaced than the intervals at which the fast paging channel is monitored. As a result of the greater spacing of the segments in the slow paging channel, power required to monitor this channel is normally less than that required to monitor the fast paging channel.

In accordance with one embodiment of the present invention different numbers of downlink control channels are monitored in different states. In such embodiments, the assignment, fast paging and slow paging channels are not monitored in all states. Rather, in the on state the greatest number of downlink control channels are monitored, fewer downlink control channels are monitored in the hold state and the lowest number of downlink control channels are monitored in the sleep state.

Fig. 7 shows a table 700 which illustrates the three exemplary base station to mobile node (downlink) control signaling channels and the corresponding four exemplary mobile node states of operation discussed above. In the table 700, a check is used to show control channels which are monitored for a given state while an X is used to indicate a control channel which is not monitored. A dashed check is used to show a control channel which may not be monitored during a portion of the time in that state but is monitored for at least a portion of the time in the state.

From Fig. 7 the first row 702 corresponds to the on state, the second row 704 corresponds to the access state, the third row 706 corresponds to the hold state and the fourth row 708 corresponds to the sleep state. Columns in the table 700 correspond to different segmented control channels. The first column 710 corresponds to the assignment channel, the second column 712 corresponds to the fast paging channel, while the third column 714 corresponds to the slow paging channel.

As can be seen from the table 700, while in the on state a mobile node 14, 16 monitors the assignment channel, fast paging control channel and slow paging control channel. For a portion of the access state, which represents a transition between the on state and either the hold state or the sleep state, the assignment and fast paging channels are monitored. The slow paging channel is monitored for the full period of time the mobile node remains in the access state. As discussed above, monitoring of the fast paging and slow paging channels requires a mobile node to be actively engaged in monitoring on a periodic, as opposed to a continuous, basis.

While in the hold state, the assignment channel is not monitored. However, the fast paging channel and slow paging channel are monitored. Accordingly, a mobile node in the hold state can be instructed to change states and/or monitor the assignment channel for traffic channel segment assignment information in a relatively short period of time.

In the sleep state, of the three control channels shown in Fig. 6, only the slow paging channel is monitored by the mobile node. Accordingly, a mobile node 14, 16 in the hold state can be instructed to change states and/or monitor the assignment channel for traffic channel

segment assignment information but such instructions may take longer to be detected, on average, than when in the hold state.

By decreasing the number of control channels that are monitored as operation proceeds from the on state to the less active sleep state, mobile node monitoring and processing resources, and thus power consumption, can be effectively controlled. Thus, the sleep state requires less mobile node resources, including power, than the hold state. Similarly, the hold state requires less mobile node resources, including power, than the on state.

Mobile node transitions from active to less active states of operation may occur in response to commands to change states received from a base station. However, in various embodiments of the invention such transitions are also initiated by mobile nodes 14, 16 in response to detecting periods of downlink control signal inactivity or reduced activity pertaining to the mobile node.

In one embodiment of the invention, activity relating to a mobile node 14, 16 on the control channel which will cease to be monitored if the mobile node reduces its state of activity by one level is used to determine when the mobile node should, on its own, switch to the lower activity level state of operation. For example, in the case of the on state, a mobile node monitors the assignment channel for signals directed to it. When failing to detect signals on the assignment channel for a preselected period of time, or a reduced message level for a period of time, the mobile node 14, 16 switches from the on state to the hold state and ceases to monitor the assignment channel.

While in the hold state, the mobile node 14, 16 monitors the fast paging channel for activity to determine, among other things, if it should switch to a lower activity state of operation, e.g., the sleep state. When failing to detect signals for a preselected period of time, or a reduced signal level for a period of time, the mobile node 14, 16 switches from the hold state to the sleep state and ceases to monitor the fast paging channel.

Using the above discussed methods, monitoring, signal processing and power resources can be conserved in a mobile node 14, 16 through the use of multiple states of operation and through the use of multiple segmented control channels. In addition, limited control resources, e.g., bandwidth used for communicating control information from a base

station to a mobile node, is used efficiently as a result of using multiple control channels, e.g., segmented control channels of the type described above.

Numerous variations on the above described methods and apparatus will be apparent to one of ordinary skill in the art in view of the above description of the invention. Such variations remain within the scope of the invention.

What is claimed is:

1. A communications method, the method comprising:

operating a first wireless terminal, at different times, in each one of three different operational states, the three different operational states including a first state, a second state and a third state,

wherein operating the first wireless terminal in the first state includes using a first amount of a control communications resource to communicate control information between said first wireless terminal and a base station and performing, using some of said first amount of a control communications resource, power control signaling at a first rate;

wherein operating the first wireless terminal in the second state includes using a second amount of the control communications resource to communicate control information between said first wireless terminal and the base station and using some of said second amount of the control communications resource to perform power control signaling at a second rate, the second rate being lower than said first rate;

wherein operating the first wireless terminal in the third state includes using less of the control communications resource than is used by the first wireless terminal in either of the first or second states, and

operating the first wireless terminal to transition from one of said first and second states to said third state, the step of transitioning from one of said first and second states to said third state including reducing the rate of power control signaling .

2. The communications method of claim 1, wherein operating the first wireless terminal in the first state to use a first amount of a control communications resource includes:

operating the first wireless terminal to perform timing control signaling operations.

3. The method of claim 1, further comprising:

operating the first wireless terminal to transition from the first state to the second state, said transitioning including reducing the rate of power control signaling performed by said first wireless terminal.

4. The method of claim 3, wherein the step of reducing the rate of power control signaling when transitioning from one of said first and second states to said third state includes ceasing to perform power control signaling.

5. The method of claim 4, further comprising: ceasing to perform timing control signaling when transitioning from said first state to said third state.

6. The method of claim 3, wherein reducing the rate of power control signaling operations includes performing transmission power control update operations at a greater interval than transmission power control update operations were performed in said first state.

7. The method of claim 6, wherein transitioning from the second state to the third state includes operating the first wireless terminal to cease performing transmission power control update operations.

8. The method of claim 2, wherein timing control update operation are performed in said first state but not in said third state, the method further comprising the step of:

operating the first wireless terminal to transition from said third state to one of said first and said second states, the step of transitioning to one of said first and second states including resuming transmission timing control update operations.

9. The method of claim 8, wherein operating the first wireless terminal to transition from said third state to one of said first state and said second state includes operating the first wireless terminal to resume transmission power control signaling operations.

10. The method of claim 9, wherein operating the first wireless terminal to resume transmission timing control signaling includes:

transmitting a request to the base station for the allocation of communications resources required to perform transmission timing control signaling.

11. The method of claim 10, wherein transmitting a request for communications resources required to perform transmission timing control signaling includes:

transmitting said request using a shared segment of a communications channel.

12. The method of claim 11, further comprising:

operating the first wireless terminal to transition from said second state to said first state, the step of transitioning from said second state to said first state including transmitting a request for a dedicated communications resource that can be used to transmit data to be communicated to said base station.

13. The method of claim 9, wherein operating the first wireless terminal to transmit a request for a dedicated communications resource that can be used to transmit data includes transmitting the resource request to a base station using a dedicated resource request uplink assigned to the first wireless terminal.

14. The method of claim 1, wherein said control communications resource is control signaling bandwidth used for communicating control signals between said base station and a plurality of wireless terminals served by said base station.

15. The method of claim 14, wherein said first state is an on state, said second state is a hold state, and said third state is a sleep state, the method further comprising:

operating the first wireless terminal to transmit and receive data as part of a communications session with another terminal during at least a portion of said on state.

16. The method of claim 15, further comprising:

operating the first wireless terminal to receive data from another terminal during at least a portion of said hold state without transmitting data at any time while operating in said hold state.

17. The method of claim 16, wherein operating said first wireless terminal in said sleep state includes controlling said wireless terminal so that data corresponding to a communications session is neither transmitted from said first wireless terminal or received by said first wireless terminal during any portion of said sleep state.

18. The method of claim 1, wherein said first state is an on state, said second state is a hold state, and said third state is a sleep state, the method further comprising:

operating a second wireless terminal, at different times, in each one of said on state, said hold state and said sleep state;

wherein operating the second wireless terminal in the on state includes using a fourth amount of control communications bandwidth to communicate control information between said second wireless terminal and said base station; and

wherein operating the second wireless terminal in the hold state includes using a fifth amount of the control communications bandwidth to communicate control information between said second wireless terminal and the base station, the fifth amount of control communications bandwidth being less than the fourth amount.

19. The method of claim 18, further comprising

operating the second wireless terminal to transmit and receive data as part of a communications session with another terminal during at least a portion of time during which said second terminal is operated in said on state.

20. The method of claim 19, further comprising:

operating the second wireless terminal to receive data from another terminal during at least a portion of the time said second wireless terminal is operated in said hold state without transmitting data at any time while operating said second wireless terminal in said hold state.

21. The method of claim 20, wherein operating said second wireless terminal in said sleep state includes controlling said second wireless terminal so that data corresponding to a communications session is neither transmitted from said second wireless terminal or received by said second wireless terminal during any portion of time said wireless terminal is operated in said sleep state.

22. The method of claim 19,

wherein said data to be transmitted as part of a communications session includes IP packets at least some of which include speech data; and

wherein said second wireless terminal transmits said IP packets to said base station using orthogonal frequency division multiplexed signals.

23. The method of claim 1, wherein operating the wireless terminal in the second state includes transmitting a control signal to initiate a state change using a dedicated uplink

communications resource said state change control signal transmission being contention free due to the use of the dedicated communications resource.

24. The method of claim 23, wherein operating the wireless terminal in said second state further includes:

transmitting a state change transition message over a shared downlink communications resource monitored by multiple wireless terminals for state transition control messages

25. The method of claim 24, wherein said first state is an on-state, said second state is a hold state and said third state is a sleep state, power control signaling being performed by said wireless terminal in the hold state at a lower rate than in said on state, zero power control signaling being performed by said wireless terminal in the sleep state.

26. The method of claim 25, wherein said wireless terminal does not perform timing control signaling in said sleep state.

27. The method of claim 1,

wherein said first state is an on state, said second state is a hold state and said third state is a sleep state,

wherein said first wireless terminal has dedicated contention free uplink and down link resource request communications resources available in said on state;

wherein said first wireless terminal has a dedicated contention free uplink resource request communications resource available in said hold state and a shared contention signaling based downlink resource request communications resource available in said hold state; and

wherein said first wireless terminal has neither a dedicated contention free uplink resource request communications resource nor a dedicated content free downlink resource request communications resource in said sleep state.

28. The method of claim 1, wherein state transitions are performed as a function of the quality of service level to be provided to said wireless terminal.

29. The method of claim 28, wherein state transitions are performed as a function of user activity.

30. The method of claim 1, wherein transitioning from said one of said first and second states is a function of user data activity.

31. The method of claim 1,

wherein operating said wireless terminal in said third state includes using a first set of communications resources;

wherein operating said wireless terminal in said second state includes using the first set of communications resources used in said in third state and a second set of communications resources; and

wherein operating said wireless terminal in said first state includes using a third set of communications resources in addition to said first and second sets of communications resources.

32. The method of claim 31, wherein said first, second and third communications resources each include communications channel segments used by said first wireless terminal.

33. The method of claim 1, wherein operating said wireless terminal in said first state and in said second state includes:

communicating timing control signals between said base station and said wireless terminal.

34. The method of claim 33, wherein timing control signals are communicated between said base station and said wireless terminal during said first state at a rate that is at least as fast as the rate at which timing control signals are communicated between said base station and said wireless terminal during said second state.

35. The method of claim 33, wherein the rate of communicating timing control signals in said first state does not exceed the rate of communicating power control signals in said first state.

36. The method of claim 35, wherein the rate of communicating timing control signals in said second state does not exceed the rate of communicating power control signals in said second state.

37. The method of claim 1, wherein operating said wireless terminal during said hold state includes:

receiving downlink signaling, other than timing control and power control signaling, using a communications channel which is monitored by multiple wireless terminals for data.

38. The method of claim 37, wherein said shared communication resource is a communications channel, wherein said downlink signal includes text information transmitted to multiple wireless terminals.

39. The method of claim 1, wherein operating said wireless terminal in said third state includes using a common uplink signaling resource to transmit a signal used to initiate a state change.

40. The method of claim 39, wherein using said common uplink signaling resource to transmit a signal used to initiate a state change includes performing contention based signaling as part of a process of transitioning from the third state to one of the first state and the second state.

41. The method of claim 1, wherein said method further includes operating a base station included in the same cell as said first wireless terminal, the base station allocating at least some downlink communications resources, used to communicate data from said base station to wireless terminals in said second state, said downlink communications resources used to communicate data being different from control communications resources used to control wireless terminal signaling activity.

42. The method of claim 1, wherein said method further includes operating a base station, included in the same cell as said wireless terminal, to allocate zero dedicated uplink data communications resources to wireless terminals in said second state, said uplink data communications resources being different from timing control and power control communications resources.

43. The method of claim 1, wherein said method further includes operating a base station, included in the same cell as said wireless terminal to control state transitions to provide different terminals in said cell with different levels of quality of service.

44. A method operating at least a first wireless terminal, the method comprising:

controlling said first wireless terminal to operate, at different times, in each one of at least three different operational states, the three different operational states including an on state, a hold state and a sleep state,

wherein operating the first wireless terminal in the on state includes communicating power control information between said first wireless terminal and a base station at a first rate;

wherein power control information is not communicated between said base station and said first wireless terminal while said first wireless terminal is operating in the sleep state;

wherein operating said wireless terminal in said hold state, includes communicating power control information between said first wireless terminal and a base station at a rate slower than said first rate, and using a dedicated uplink communications resource to transmit information to said base station, said dedicated uplink communications resource being in addition to any power control signaling and timing control signaling communication resource used by said base station during said hold state; and

transitioning from one of said three states to another one of said three states in response to a change is user activity.

45. The method of claim 44 wherein said information transmitted using said dedicated uplink communications resource during said hold state includes a signal used to initiate a state transition, said signal being transmitted contention free due to the use of at least one uplink communications resource segment dedicated to said wireless terminal for transmission of said signal.

46. A communications device including:

control means used to control the communications device to:

i) operate in an on state in which the communications device uses a first amount of control signaling resources for the communication of control information between said communications device and a base station operating in said on state including performing power control signaling at a first rate;

ii) transition the communications device from the on state to a hold state in which the first communications device uses less control signaling resources for the communication of control information between said device and a base station

than is used by the communications device in the on state said transitioning including reducing the rate of power control signaling; and

iii) transition the communications device from said hold-state to a sleep state in which the communications device uses less control signaling resources than the communications device uses in said hold state, said transition from the on state to the hold state including reducing the rate of power control signaling.

47. The communications device of claim 46, wherein said control means controls the communications device not to perform power control signaling during said sleep state.

48. The communications device of claim 47, wherein said control means controls the communications device to perform transmission timing control signaling operations, transmission power control signaling operations, and transmit data, during said on state.

49. The communications device of claim 48, wherein said control means controls the communications device, as part of transitioning from said hold-state to said sleep-state, to cease performing transmission timing control update operations.

50. The communications device of claim 48, further comprising:

means for controlling the communications device to transition from said sleep-state to one of said on-state and said hold-state, wherein transitioning to one of said on-state and said hold-state includes resuming transmission timing control update operations.

51. The communications device of claim 46, further comprising:

means for controlling the communications device to transition from said hold-state to said on-state, wherein transitioning from said hold-state to said-on state includes transmitting a request for a dedicated communications resource that can be used to transmit data to be communicated to said base station, said request being transmitted contention free by using a dedicated portion of resource request communications channel.

52. A communications system comprising:

a base station, said base station controlling the allocation of control signaling resources and data transmission resources to a plurality of nodes serviced by said base station, said base station controlling a first subset of said plurality of nodes to operate in an on state

wherein nodes in said first subset are allocated data communication resources to transmit data and control signaling resources to perform a first level of control signaling, said base station further controlling a second subset of said plurality of nodes to operate in an hold state wherein nodes in said second subset are allocated control signaling resources to perform a second level of control signaling which is less than said first level of control signaling; and

said base station further controlling a third subset of said plurality of nodes to operate in a sleep state wherein nodes in said third subset are allocated less control signaling resources than nodes in either said first subset or said second subset, said base station allocating more power control signaling resources to a node in said on state than to a node in a hold state, said base station allocating more power control signaling resources to a node in said hold state than a node in said sleep state.

53. The communications system of claim 52, wherein the system includes said plurality of nodes, the first subset of nodes includes means for performing transmission timing control signaling operations while in said on state.

54. The communications system of claim 53, wherein the second subset of nodes includes means for performing transmission timing control signaling operations and reduced rate transmission power control signaling operations while in said hold state.

55. The communication system of claim 53, wherein said second set of nodes includes means for halting transmission power control update operations when transitioning into said hold state from said on state.

56. The communication system of claim 53, wherein said third subset of nodes includes means for terminating the transmission of timing control signals when transitioning into said sleep state from said hold state.

57. The system of claim 52, further comprising said plurality of nodes, the third subset of nodes including more communications devices than said second subset of nodes.

58. The system of claim 57, wherein the second subset of nodes includes more nodes than said first subset of nodes.

59. The system of claim 58, wherein said third subset of nodes do not perform transmission power control signaling operations.

60. The system of claim 59, wherein said second subset of nodes do not perform transmission power control signaling operations.

61. The system of claim 59, wherein said second subset of nodes perform power control signaling operations at a rate which is lower than the rate at which nodes in said first subset perform transmission power control signaling operations.

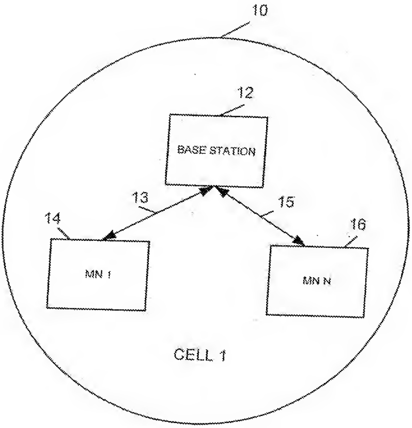


Fig. 1

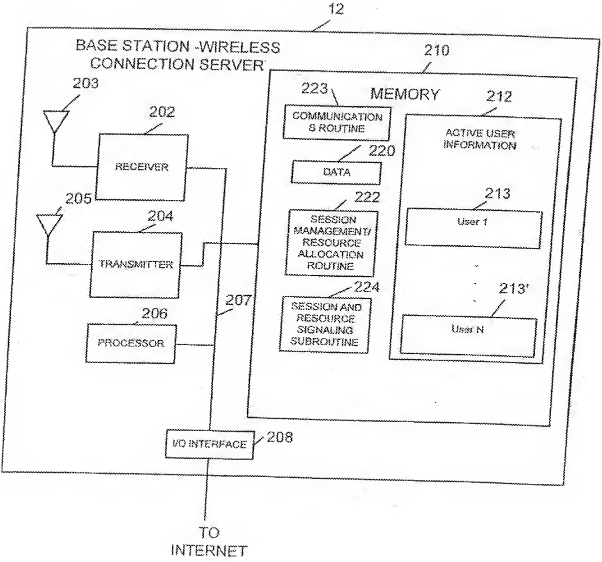

12
BASE STATION -WIRELESS CONNECTION SERVER
210
203
223
MEMORY
212
202
COMMUNICATIONS ROUTINE
ACTIVE USER INFORMATION
RECEIVER
220
DATA
205
222
213
204
SESSION MANAGEMENT/ RESOURCE ALLOCATION ROUTINE
User 1
TRANSMITTER
224
213'
207
206
SESSION AND RESOURCE SIGNALING SUBROUTINE
User N
PROCESSOR
I/O INTERFACE
208
TO INTERNET

Fig. 2

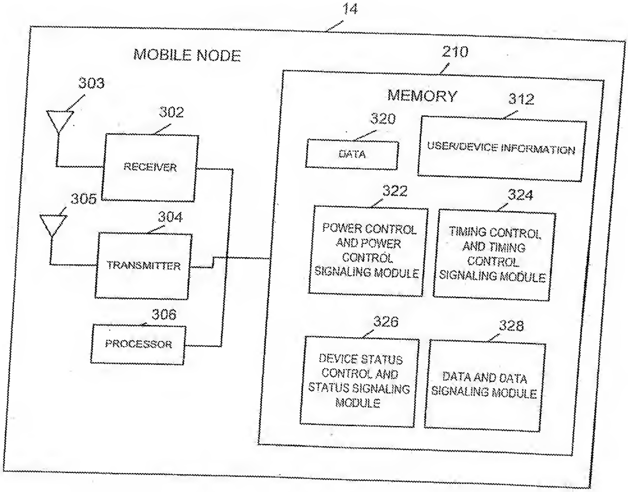

14
MOBILE NODE
303
302
RECEIVER
305
304
TRANSMITTER
306
PROCESSOR
210
MEMORY
320
DATA
312
USER/DEVICE INFORMATION
322
POWER CONTROL AND POWER CONTROL SIGNALING MODULE
324
TIMING CONTROL AND TIMING CONTROL SIGNALING MODULE
326
DEVICE STATUS CONTROL AND STATUS SIGNALING MODULE
328
DATA AND DATA SIGNALING MODULE

Fig. 3

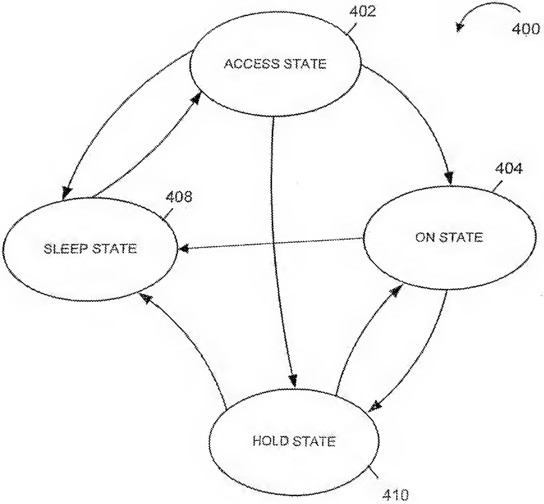

400
402
ACCESS STATE
404
ON STATE
408
SLEEP STATE
HOLD STATE
410

Fig. 4

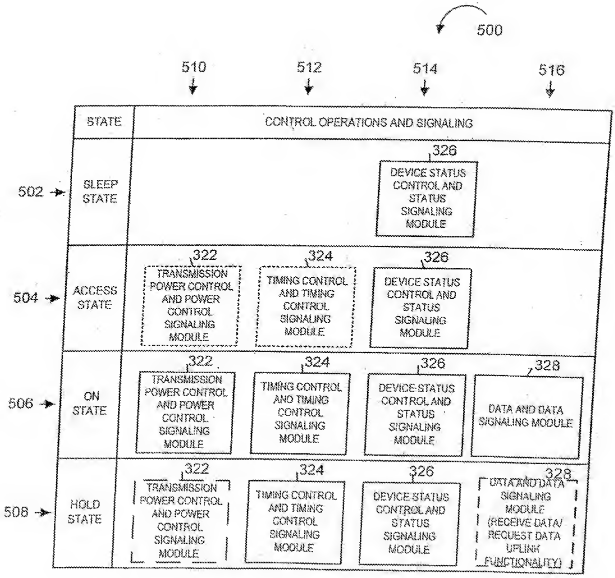

500
510
512
514
516
STATE
CONTROL OPERATIONS AND SIGNALING
502
SLEEP STATE
326
DEVICE STATUS CONTROL AND STATUS SIGNALING MODULE
504
ACCESS STATE
322
TRANSMISSION POWER CONTROL AND POWER CONTROL SIGNALING MODULE
324
TIMING CONTROL AND TIMING CONTROL SIGNALING MODULE
326
DEVICE STATUS CONTROL AND STATUS SIGNALING MODULE
506
ON STATE
322
TRANSMISSION POWER CONTROL AND POWER CONTROL SIGNALING MODULE
324
TIMING CONTROL AND TIMING CONTROL SIGNALING MODULE
326
DEVICE STATUS CONTROL AND STATUS SIGNALING MODULE
328
DATA AND DATA SIGNALING MODULE
508
HOLD STATE
322
TRANSMISSION POWER CONTROL AND POWER CONTROL SIGNALING MODULE
324
TIMING CONTROL AND TIMING CONTROL SIGNALING MODULE
326
DEVICE STATUS CONTROL AND STATUS SIGNALING MODULE
328
DATA AND DATA SIGNALING MODULE (RECEIVE DATA/ REQUEST DATA UPLINK FUNCTIONALITY)

Fig. 5

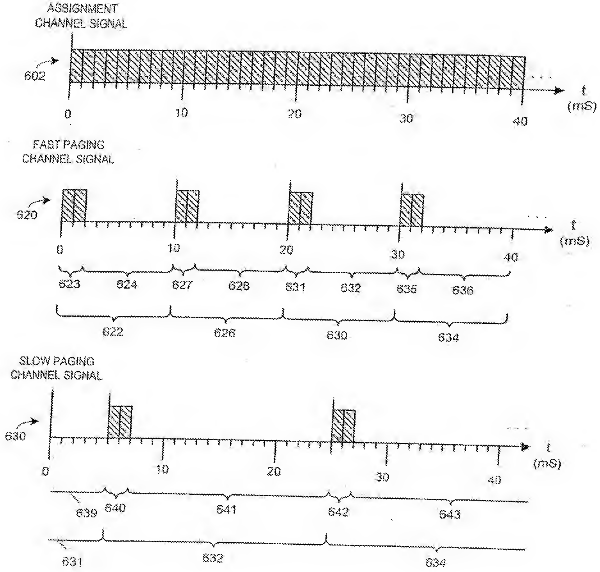

ASSIGNMENT CHANNEL SIGNAL
602
0
10
20
30
40
t (mS)
FAST PAGING CHANNEL SIGNAL
620
0
10
20
30
40
t (mS)
623
624
627
628
631
632
635
636
622
626
630
634
SLOW PAGING CHANNEL SIGNAL
630
0
10
20
30
40
t (mS)
639
640
641
642
643
631
632
634

Fig. 6

700

	STATE	ASSIGNMENT CHANNEL (710)	FAST PAGING CHANNEL (712)	SLOW PAGING CHANNEL (714)
702	ON STATE	✓	✓	✓
704	ACCESS STATE	✓	✓	✓
706	HOLD STATE	✗	✓	✓
708	SLEEP STATE	✗	✗	✓

Fig. 7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US03/24889

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : H04Q 7/20
US CL : 455/343.2, 522, 574; 340/7.32; 370/311

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 455/343.2-343.4, 522, 574; 340/7.32-7.36; 370/311

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y --- A	US 6,275,712 B1 (GRAY et al) 14 August 2001, column 1, lines 45-65.	1-9, 14-17, 23-26, 28-50, 52-61 --- 10-13, 18-22, 27, 51
Y --- A	US 6,334,047 B1 (ANDERSSON et al) 25 December 2001, column 9, lines 11-24.	1-9, 14-17, 23-26, 28-50, 52-61 --- 10-13, 18-22, 27, 51

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
12 November 2003 (12.11.2003)

Date of mailing of the international search report
05 DEC 2003

Name and mailing address of the ISA/US
Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
Facsimile No. (703)305-3230

Authorized officer
LEE NGUYEN
Eugenia Logan
Telephone No. 703-305-5249

INTERNATIONAL SEARCH REPORT

PCT/US03/24889

Continuation of B. FIELDS SEARCHED Item 3:
EAST
search term: active, hold, sleep state

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) **World Intellectual Property Organization**
International Bureau

(43) **International Publication Date**
4 November 2004 (04.11.2004)

PCT

(10) **International Publication Number**
WO 2004/095854 A2

(51) **International Patent Classification**[7]: **H04Q**

(21) **International Application Number:** PCT/US2004/008272

(22) **International Filing Date:** 17 March 2004 (17.03.2004)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**
10/397,493 26 March 2003 (26.03.2003) US

(71) **Applicant** *(for all designated States except US)*: **MOTOROLA INC.** [US/US]; 1303 EAST ALGONQUIN ROAD, SCHAUMBURG, IL 60196 (US).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **AERRABOTU, Naveen** [IN/US]; 7943 Dada Drive, Gurnee, IL 60031 (US). **BINZEL, Charles P.** [US/US]; 17001 60th Street, Bristol, WI 53104 (US). **SRINIVASAN, Bharat** [IN/US]; 4245 N. Bloomington, Arlington Heights, IL 60004 (US).

(74) **Agents: LOPPNOW, Matthew C.** et al.; 600 North US Highway 45, Libertyville, IL 60048 (US).

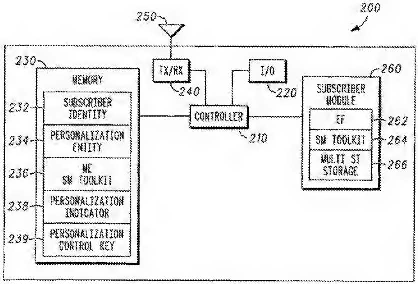
(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) **Title:** METHOD AND APPARATUS FOR MULTIPLE SUBSCRIBER IDENTITIES IN A MOBILE COMMUNICATION DEVICE



(57) **Abstract:** A method and apparatus for multiple subscriber identities in a single subscriber module in a mobile communication device. The method can include storing the plurality of subscriber identities on the single mobile communication device subscriber module and personalizing the mobile communication device to each of the plurality of subscriber identities. The mobile communication device can issue a single command to retrieve all subscriber identities from the single mobile communication device subscriber module and store the plurality of subscriber identities from the single mobile communication device subscriber module to a memory of the mobile communication device. The plurality of subscriber identities can be stored in a single elementary file or another location on the single mobile communication device subscriber module.

WO 2004/095854 A2

METHOD AND APPARATUS FOR MULTIPLE SUBSCRIBER IDENTITIES IN A MOBILE COMMUNICATION DEVICE

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is related to the application entitled "Method and Apparatus for Updating a Subscriber Identity in a Mobile Communication Device," Motorola case number CS22214RL, filed on even date herewith and commonly assigned to the assignee of the present application.

BACKGROUND OF THE INVENTION

1. Field of Invention

The present invention is directed to a method and apparatus for multiple subscriber identities in a mobile communication device. In particular, the present invention is directed storing and accessing multiple subscriber identities on a subscriber module for a mobile communication device.

2. Description of Related Art

Presently a mobile communication device may be personalized to a particular security module for security purposes. Such a security module may be a Subscriber Identity Module (SIM), a User Services Identity Module (USIM), or any other security module. The security module can contain a single subscriber identity such as a code group, an International Mobile Subscriber identity (IMSI), or any other subscriber identity. The personalization is an anti-theft feature. When a mobile communication device is personalized to a particular security module, it can refuse to operate with another security module. Thus, if the mobile communication device is stolen, the thief cannot use the mobile communication device with another security module. While this does not prevent the mobile communication device from being stolen, it does make the mobile communication device less attractive to the thief.

The mobile communication device can be personalized by storing the single subscriber identity of the current subscriber module in the mobile communication device and setting a personalization indicator to "on." Then, whenever a subscriber

module is inserted or the mobile communication device is powered up with a subscriber module in place, the single subscriber identity is read from the subscriber module and checked against the stored subscriber identity. If there is no match, access to the mobile communication device functions is blocked and the mobile communication device may only allow emergency calls to be placed from the mobile communication device.

Unfortunately, typically a mobile communication device may only be personalized to a single subscriber identity. Also, a subscriber module for a mobile communication device cannot store easily accessible multiple subscriber identities. Additionally, a mobile communication device cannot easily access multiple subscriber identities on a single subscriber module. Furthermore, a mobile communication device cannot personalize to multiple subscriber identities on a single subscriber module. Thus, there is a need for a method and apparatus for multiple subscriber identities in a mobile communication device subscriber module.

BRIEF DESCRIPTION OF THE DRAWINGS

The preferred embodiments of the present invention will be described with reference to the following figures, wherein like numerals designate like elements, and wherein:

Fig. 1 is an exemplary block diagram of a system according to a preferred embodiment;

Fig. 2 is an exemplary block diagram of a mobile communication device including a subscriber module according to a preferred embodiment;

Fig. 3 is an exemplary illustration of a mobile communication device subscriber module according to another embodiment;

Fig. 4 is an exemplary flowchart outlining the operation of a subscriber module and a mobile communication device according to a preferred embodiment;

Fig. 5 is an exemplary message sequence chart outlining the operation of a mobile communication device according to another embodiment;

Fig. 6 is an exemplary flowchart outlining the operation of the mobile communication device according to another embodiment; and

Fig. 7 is an exemplary message sequence chart outlining operation of the mobile communication device according to another embodiment.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

The present disclosure provides an apparatus and method for updating a subscriber identity in a mobile communication device. For example, the present disclosure provides a method in a mobile communication device having a single mobile communication device subscriber module including a plurality of subscriber identities. The method can include storing the plurality of subscriber identities on the single mobile communication device subscriber module and personalizing the mobile communication device to each of the plurality of subscriber identities. The method can also include operating the mobile communication device on a first network using a first subscriber identity, detecting a change of network coverage to a second network, and switching an operational subscriber identity from the first subscriber identity to a second subscriber identity based on the change of network coverage. Detecting a change can include detecting a change from a first service cell to a second service cell. The first service cell may provide billing based on first subscriber and the second service cell may provide billing based on second subscriber.

Personalizing the mobile communication device can include issuing a select command to the single mobile communication device subscriber module, the select command selecting a subscriber identity elementary file on the single mobile communication device subscriber module, the subscriber identity elementary file containing the plurality of subscriber identities, sending a read command to the single mobile communication device subscriber module, receiving the plurality of subscriber identities from the single mobile communication device subscriber module in response to sending the read command, storing the plurality of subscriber identities in a memory of the mobile communication device, and setting a personalization indicator to on.

Personalizing the mobile communication device may also include reading a first subscriber identity from a subscriber identity elementary file, updating the subscriber identity elementary file with a second subscriber identity, and reading the second subscriber identity from the subscriber identity elementary file. Personalizing the mobile communication device may additionally include issuing a single command

to retrieve all subscriber identities from the single mobile communication device subscriber module, storing the plurality of subscriber identities from the single mobile communication device subscriber module to a memory of the mobile communication device, and setting a personalization indicator to on.

The plurality of subscriber identities can be stored in a single elementary file or in other locations on the single mobile communication device subscriber module. The single mobile communication device subscriber module can be a Subscriber Identity Module and the subscriber identity can be an International Mobile Subscriber Identity.

According to another embodiment, the present disclosure can provide a method in a mobile communication device including a single subscriber module having a plurality of subscriber identities. The method can include issuing a select command to the single subscriber module, the select command selecting a subscriber identity elementary file on the single subscriber module, the subscriber identity elementary file containing the plurality of subscriber identities. The method can also include receiving a response from the subscriber module in response to issuing the select command, the response including a file size of the subscriber identity elementary file. The method can additionally include sending a read command to the single subscriber module and receiving the plurality of subscriber identities from the single subscriber module in response to sending the read command. The read command can include an offset parameter indicating an offset in the subscriber identity elementary file, and a length parameter indicating a length of the data to be read.

The method can further include storing the plurality of subscriber identities in a memory of the mobile communication device and setting a personalization indicator in the mobile communication device to on. The method can additionally include operating the mobile communication device on a first network using a first subscriber identity, detecting a change of network coverage to a second network, and switching an operational subscriber identity from the first subscriber identity to a second subscriber identity based on the change of network coverage, based on a location

status message from the network, or any other service provider or card manufacturer algorithm.

The method can also include reading a subscriber identity from the single subscriber module, comparing the subscriber identity with the plurality of subscriber identities stored in the mobile communication device, and blocking use of selected features of the mobile communication device if the subscriber module subscriber identity does not match one of the plurality of subscriber identities stored in the mobile communication device. The subscriber module may be a Subscriber Identity Module and the subscriber identity may be an International Mobile Subscriber Identity.

According to another embodiment, the present disclosure can provide a mobile communication device subscriber module. The mobile communication device subscriber module can include a controller configured to control the operations of the mobile communication device subscriber module, an input and output contact point coupled to the controller, and a supply voltage contact point coupled to the controller. The mobile communication device subscriber module can also include a memory including a multiple subscriber identity elementary file. The multiple subscriber identity elementary file can have a body including a plurality of subscriber identity locations and a plurality of subscriber identities, each subscriber identity consisting of eight bytes. Each of the plurality of subscriber identity locations can include at least a subscriber identity of the plurality of subscriber identities. The multiple subscriber identity elementary file can include a mandatory first subscriber identity of eight bytes. The subscriber module can be a Subscriber Identity Module and the subscriber identity can be an International Mobile Subscriber Identity. The controller can be configured to operate the mobile communication device subscriber module on a first network using a first subscriber identity, detect a change of network coverage to a second network, and switch an operational subscriber identity from the first subscriber identity to a second subscriber identity based on the change of network coverage. The controller can also be configured to personalize a mobile communication device to the plurality of subscriber identities. The controller can be further configured to receive a select command from a mobile communication device, the select command selecting

the multiple subscriber identity elementary file, accept a read command from the mobile communication device, and send the plurality of subscriber identities from the subscriber module in response to accepting the read command. The memory can also include a single subscriber identity elementary file comprising a body including a single subscriber identity.

According to another embodiment, the present disclosure can provide a method in a mobile communication device including a plurality of subscriber identities on a single mobile communication device subscriber module. The method can include storing the plurality of subscriber identities on the single mobile communication device subscriber module and issuing a retrieve command for retrieving all of the plurality of subscriber identities on the single mobile communication subscriber module. The method can also include receiving a subscriber identity amount indicator, the subscriber identity amount indicator indicating a number of subscriber identities located on the single mobile communication subscriber module. The method can additionally include receiving all of the plurality of subscriber identities from the single mobile communication subscriber module in response to sending the read command, and storing all of the plurality of subscriber identities to a memory of the mobile communication device. The method can further include personalizing the mobile communication device to each of the plurality of subscriber identities by setting a personalization indicator to on. The method can also include switching an operational subscriber identity from a first subscriber identity to a second subscriber identity. For example, the method can include operating the mobile communication device on a first network using a first subscriber identity, detecting a change of network coverage to a second network, and switching an operational subscriber identity from the first subscriber identity to a second subscriber identity based on the change of network coverage. The plurality of subscriber identities may be stored in a single elementary file or in any other location on the single mobile communication device subscriber module.

Among other benefits, the present invention can allow for storing and accessing multiple subscriber identities on a subscriber module for a mobile communication device. The present invention can additionally provide for

personalizing a mobile communication device to multiple subscriber identities stored on a single subscriber module. The present invention can also provide for easily and readily accessing multiple subscriber identities on a single subscriber module. The present invention can further provide for an elementary file for storing multiple subscriber identities on a single subscriber module. The present invention can additionally provide for changing between multiple subscriber identities during operation of a mobile communication device. These and further benefits will become more apparent with reference to the Figures and the descriptions of the preferred embodiments.

Fig. 1 is an exemplary block diagram of a system 100 according to a preferred embodiment. The system 100 can include a network controller 110, a network 120, and one or more mobile communication devices 130 and 140. The mobile communication devices 130 and 140 may be mobile equipment such as wireless telephones, cellular telephones, personal digital assistants, or any other devices that are capable of sending and receiving voice and data signals over a wireless network.

In an exemplary embodiment the network controller 110 is connected to the network 120. The network controller 110 may be included in a base transceiver station, a service center, or any other device on the network 120. The network 120 may include any type of network that is capable of sending and receiving communication signals. For example, the network 120 may include a data network, such as the Internet, an Intranet, a local area network (LAN), a wide area network (WAN), a cable network, and other like communication systems. The network 120 may also include a telecommunications network, such as a local telephone network, long distance telephone network, cellular telephone network, satellite communications network, cable television network and other like communications systems. Furthermore, the network 120 may include more than one network and may include a plurality of different types of networks. Thus, the network 120 may include a plurality of data networks, a plurality of telecommunications networks, a combination of data and telecommunications networks and other like communication systems. Preferably, the network 120 is a wireless network.

In operation, the network controller 110 can control operations on the network 120. The mobile communication devices 130 and 140 can transmit and receive wireless signals to and from the network 120. For example, the mobile communication device 130 can connect a voice call with the mobile communication device 140. Thus, users of the mobile communication devices 130 and 140 can audibly communicate with each other. Also, the mobile communication device 130 can connect a data call with the mobile communication device 140. Thus, users of the mobile communication devices 130 and 140 can send and receive data to and from each other and the network 120.

Fig. 2 is an exemplary block diagram of a mobile communication device 200, such as the mobile communication device 130, according to a preferred embodiment. The mobile communication device 200 can include a controller 210, input and output circuitry 220, a memory 230, a transceiver 240, an antenna 250 and a removable subscriber module 260. The subscriber module 260 may be a subscriber identity module such as a user services identity module, a removable user identity module, or any other subscriber module. The subscriber module 260 can include at least one elementary file 262, a subscriber module toolkit 264, and a multiple subscriber identity storage 266. The elementary file 262 can contain one subscriber identity. The multiple subscriber identity storage 266 may be a multiple subscriber identity elementary file, a hidden location for storing multiple subscriber identities, or any other storage location or locations for storing multiple subscriber identities. The subscriber module toolkit 264 can interact with the mobile communication device 200 to perform various functions with the mobile communication device 200.

The memory 230 may be a random access memory, a read only memory, an optical memory, or any other memory. The memory 230 can include subscriber identity storage 232, a personalization entity 234, a mobile equipment or mobile communication device subscriber module toolkit 236, a personalization indicator 238, and a personalization control key 239. The personalization entity 234 and the mobile communication device subscriber module toolkit 236 may also reside on the controller 210, as independent software or hardware modules, or in any other format on the mobile communication device 200. The subscriber identity storage 232 can contain a

subscriber identity that the mobile communication device 200 is currently personalized to. The subscriber identity storage 232 may also contain a personalization list containing one or multiple subscriber identities. The personalization entity 234 can perform functions to personalize the mobile communication device 200 to a subscriber identity stored in the subscriber identity storage 232. The mobile communication device subscriber module toolkit 236 can control operations and communications with the subscriber module 260. The personalization indicator 238 can be set to "on" to indicate the mobile communication device 200 is personalized to a subscriber identity of a subscriber module. The personalization control key 239 controls the personalization of the mobile communication device 200. For example, the personalization control key 239 can be selected by a user to allow for de-personalization of the mobile communication device 200.

The transceiver 240 may include a transmitter and/or a receiver. The input and output circuitry 220 can include a microphone, a display, a speaker, a user input such as a keypad and buttons, or any other input and output circuitry.

In operation, the input and output circuitry 220 can accept various forms of input and output signals. For example, the input and output circuitry 220 can receive and output audio signals and data signals. The memory 230 can store data and software used in the mobile communication device 200. The transceiver 240 can transmit and/or receive data over a wireless network such as network 120. The controller 210 can control the operation of the mobile communication device 200.

When the mobile communication device 200 is personalized to a particular subscriber module or at least one subscriber identity, it can refuse to operate with any other subscriber modules or subscriber identities. The mobile communication device 200 can be personalized by storing at least one subscriber identity, such as a subscriber module code group, of the relevant subscriber module 260 in the subscriber identity storage 232 and setting the personalization indicator 238 to on. Whenever a subscriber module is inserted or the mobile communication device 200 is powered up with a subscriber module already in place, the at least one subscriber identity is read from the subscriber module and checked against the at least one subscriber identity

stored in the subscriber identity storage 232. If there is no match, the mobile communication device 200 can go into an emergency calls only mode where only emergency calls can be placed from the mobile communication device 200.

As mentioned, the personalization control key 239 controls the personalization feature. This key can be selected by a user at personalization and can be later entered into the mobile communication device 200 to depersonalize the mobile communication device 200. The controller 210 can also support multiple instances of subscriber module personalization. For example, a subscribe identity can be read from a subscriber module and checked against a list of subscriber identities stored in the subscriber identity storage 232.

Personalizing the mobile communication device 200 may alternately include reading a first subscriber identity from the subscriber identity elementary file 262. The first subscriber identity can be stored in the memory 230. The subscriber identity elementary file 262 may then be updated with a second subscriber identity, for example, from the multiple subscriber identity storage 266. The second subscriber identity may then be read from the subscriber identity elementary file 262 and stored in the memory 230. The personalization indicator 238 may then be set to "on."

A personalization check is performed whenever a subscriber module 260 is inserted into the mobile communication device 200 or whenever the mobile communication device 200 is powered up with a subscriber module 260 already in place. When more than one personalization is active in the mobile communication device 200, normal mode of operation can include performing any outstanding personalization checks. To perform a personalization check, the controller 210 first checks whether the mobile communication device 200 is personalized by checking the personalization indicator 238. If the personalization indicator 238 is set to "off," the personalization check can be stopped and the mobile communication device 200 can go into normal mode of operation and omit the remaining steps of the check. If the personalization indicator 238 is set to "on," the controller 210 continues the check. The controller 210 can then read the subscriber identity from the subscribe module 260. For example, the controller 210 can read the subscriber identity from the elementary file 262. The controller 210 can then perform the personalization check by

checking the read subscriber identity against a subscriber identity or list of subscriber identities stored in the subscriber identity storage 232. If no match is found, the controller 210 can display an appropriate message on a display such as "Insert correct subscriber module" and can then go into an emergency calls only mode. Alternately, a user may be prompted to enter a special de-personalization code to de-personalize the mobile communication device 200 and allow for normal operation. Otherwise, if a match is found, the mobile communication device 200 can go into normal mode of operation.

According to a related embodiment, the personalization of a mobile communication device 200 results in the personalization control key 239 being set, the personalization indicator 238 being set to "on," and the storage, in the subscriber identity storage 232, of at least one subscriber identity to which the mobile communication device 200 is personalized. To personalize the mobile communication device 200, first, the relevant subscriber identity is entered into the subscriber identity storage 232 either by reading the subscriber identity from the subscriber module 260, or by any other process such as one defined by a manufacturer. Second, the controller 210 can perform any necessary setup and necessary pre-personalization checks that may be useful for personalization. If the necessary checks are correct, the subscriber identity can be stored in the subscriber identity storage 232. If the checks fail, the personalization process can be terminated. Third, to personalize the mobile communication device 200 to more than one subscriber module, the above steps can be repeated. Fourth, the personalization control key 239 can be stored. A single personalization control key 239 can be used for both single and multiple subscriber module personalization. Finally, the personalization indicator 238 can be set to "on."

To de-personalize the mobile communication device 200, the correct personalization control key 239 can be entered. The subscriber module 260 may or may not be present for de-personalization. If the subscriber module 260 is present, de-personalization may be offered regardless of whether any useful subscriber module personalization checks pass or fail. De-personalization can be provided by entry on a keypad of the input and output circuitry 220. Other de-personalization methods may

be used. To de-personalize the mobile communication device 200, a user can enter the personalization control key. If the entered personalization control key is the same as the one stored in the memory 230, the personalization indicator 238 is set to "off." If the entered and stored personalization control key values differ, the de-personalization process can stop and the mobile communication device 200 can remain personalized.

Fig. 3 is an exemplary illustration of a mobile communication device subscriber module 300 according to another embodiment. The subscriber module 300 can include a frame 310 and a subscriber module controller 320 configured to control the operations of the subscriber module 300. The subscriber module 300 can also include an input and output contact point 340 coupled to the subscriber module controller 320, a supply voltage contact point 350 coupled to the subscriber module controller 320, and a memory 330 coupled to the subscriber module controller 320. The input and output contract point 340 may include multiple contacts for sending and receiving communications. The memory 330 can include a multiple subscriber identity storage 334. The multiple subscriber identity storage 334 may be a multiple subscriber identity elementary file, a storage location, storage locations, or any other storage for multiple subscriber identities. For example, the multiple subscriber identity elementary file can include a body having a plurality of subscriber identity locations and a plurality of subscriber identities, each of the plurality of subscriber identity locations comprising at least a subscriber identity of the plurality of subscriber identities. The memory 330 can also include a single subscriber identity elementary file 332. The single subscriber identity elementary file 332 can have a body including a single subscriber identity. According to one embodiment, only the single subscriber identity elementary file 332 may be visible to applications on a mobile communication device. This may assist in preventing hackers from viewing or altering the contents of the multiple identity storage 334.

A subscriber identity may consist of eight bytes. Also, a multiple subscriber identity elementary file can include a mandatory first subscriber identity of eight bytes. The subscriber module 300 can be a Subscriber Identity Module and a subscriber identity can be an International Mobile Subscriber Identity.

The subscriber module controller 320 can be further configured to operate the mobile communication device subscriber module 300 on a first network using a first subscriber identity, detect a change of network coverage to a second network, and switch an operational subscriber identity from the first subscriber identity to a second subscriber identity based on the change of network coverage. The subscriber module controller 320 can also be configured to personalize a mobile communication device to the plurality of subscriber identities. The subscriber module controller 320 can additionally be configured to receive a select command from a mobile communication device, the select command selecting the multiple subscriber identity storage 334, accept a read command from a mobile communication device, and send the plurality of subscriber identities from the subscriber module 300 to the mobile communication device in response to accepting the read command.

Fig. 4 is an exemplary flowchart 400 outlining the operation of a subscriber module 260 and a mobile communication device 200 according to a preferred embodiment. In step 410, the flowchart begins. In step 420, multiple subscriber identities are stored on the subscriber module 262. For example, multiple subscriber identities may be stored in an elementary file, or in any other useful location on the subscriber module 260. In step 430, the mobile communication device 200 is personalized to the multiple subscriber identities. In step 440, the flowchart ends.

Fig. 5 is an exemplary message sequence chart 500 outlining personalization of the mobile communication device 200 according to another embodiment. In step 510, the mobile communication device 200 can issue a select command to the single mobile communication device subscriber module 260. The select command can select a subscriber identity elementary file on the single mobile communication device subscriber module 260. The subscriber identity elementary file can contain the plurality of subscriber identities. The mobile communication device 200 may receive a response from the subscriber module 260 in response to issuing the select command. For example, the response can include a file size of the subscriber identity elementary file, a file identification of the subscriber identity elementary file, type of file information, access condition information, or any other useful information.

In step 520, the mobile communication device 200 can send a read command to the single mobile communication device subscriber module 260. The read command can include an offset parameter indicating an offset in an elementary file, and a length parameter indicating a length of the data to be read.

In step 530, the mobile communication device 200 can receive the plurality of subscriber identities from the single mobile communication device subscriber module 260 in response to sending the read command. In step 540, the mobile communication device 200 can store the plurality of subscriber identities in a memory 230 of the mobile communication device 200. In step 550, the mobile communication device 200 can set a personalization indicator 238 to "on."

Fig. 6 is an exemplary flowchart 600 outlining the operation of the mobile communication device 200 according to another embodiment. In step 610, the flowchart begins. In step 620, the mobile communication device 200 operates on a network, such as network 120. For example, the network may be a local area network, a wide area network, a single cell of multiple cells in a network, or any other network. In step 630, the mobile communication device 200 can detect a change in network coverage to another network. For example, the mobile communication device 200 may move into another cell. As another example, the mobile communication device 200 may move from network coverage provided by one service provider to network coverage provided by another service provider. As another example, the mobile communication device 200 may switch modes of operation based on network coverage. As another example, the mobile communication device 200 may detect a position change using a positioning detection device such as a global positioning system device. As another example, the mobile communication device 200 may switch between networks that provide billing to different subscribers. For example, an employer may pay for mobile device service while an employee is at work, while the employee may pay for service elsewhere as determined by network coverage, a global positioning system device, or otherwise. If there is no change in network coverage, the mobile communication device 200 continues operation on the network in step 620. If there is a change in network coverage, in step 640 the subscriber module 260 on the mobile communication device 200 can switch an

operational subscriber identity from an existing operational subscriber identity to another subscriber identity. Thus, the subscriber module 260 can change subscriber identities depending on network coverage. For example, the subscriber module 260 can keep the same subscriber identity in certain networks and change the subscriber identity for other networks.

Fig. 7 is an exemplary message sequence chart 700 outlining operation of the mobile communication device 200 according to another embodiment. A plurality of subscriber identities may be stored on the subscriber module 260 in a single file or in any other locations. In step 710, a single multiple subscriber identity retrieve command can be issued to retrieve all subscriber identities from the single mobile communication device subscriber module 260. A subscriber identity amount indicator may be received from the subscriber module 260, the subscriber identity amount indicator indicating a number of subscriber identities located on the single mobile communication subscriber module. In step 720, the mobile communication device 200 can receive a plurality of subscriber identities from the subscriber module 260 in response to sending the read command. In step 730, the plurality of subscriber identities from the single mobile communication device subscriber module 260 can be stored 730 to a memory 230 of the mobile communication device 200. In step 740, if personalization is desired, a personalization indicator in the mobile communication device 200 can be set to "on.

After receiving all of the subscriber identities, the mobile communication device 200 can switch an operational subscriber identity from a first subscriber identity to a second subscriber identity. For example the mobile communication device can operate on a first network using a first subscriber identity, detect a change of network coverage to a second network, and switch an operational subscriber identity from the first subscriber identity to a second subscriber identity based on the change of network coverage.

The method of this invention is preferably implemented on a programmed processor. However, the network controller 110, the controller 210, and/or the subscriber module controller 320 may also be implemented on a general purpose or special purpose computer, a programmed microprocessor or microcontroller and

peripheral integrated circuit elements, an ASIC or other integrated circuit, a hardware electronic or logic circuit such as a discrete element circuit, a programmable logic device such as a PLD, PLA, FPGA or PAL, or the like. In general, any device on which resides a finite state machine capable of implementing the flowcharts shown in the Figures may be used to implement the processor functions of this invention.

While this invention has been described with specific embodiments thereof, it is evident that many alternatives, modifications, and variations will be apparent to those skilled in the art. For example, various components of the embodiments may be interchanged, added, or substituted in the other embodiments. Accordingly, the preferred embodiments of the invention as set forth herein are intended to be illustrative, not limiting. Various changes may be made without departing from the spirit and scope of the invention.

WHAT IS CLAIMED IS:

1. A method in a mobile communication device having a single mobile communication device subscriber module including a plurality of subscriber identities, comprising:

storing the plurality of subscriber identities on the single mobile communication device subscriber module; and

personalizing the mobile communication device to each of the plurality of subscriber identities.

2. The method according to claim 1, further comprising:

operating the mobile communication device on a first network using a first subscriber identity;

detecting a change of network coverage to a second network; and

switching an operational subscriber identity from the first subscriber identity to a second subscriber identity based on the change of network coverage.

3. The method according to claim 2, wherein detecting a change comprises detecting a change from a first service cell to a second service cell.

4. The method according to claim 3, wherein the first service cell provides billing based on first subscriber identity and the second service cell provides billing based on second subscriber identity.

5. The method according to claim 1, wherein personalizing the mobile communication device comprises:

issuing a select command to the single mobile communication device subscriber module, the select command selecting a subscriber identity elementary file on the single mobile communication device subscriber module, the subscriber identity elementary file containing the plurality of subscriber identities;

sending a read command to the single mobile communication device subscriber module;

receiving the plurality of subscriber identities from the single mobile communication device subscriber module in response to sending the read command;

storing the plurality of subscriber identities in a memory of the mobile communication device; and

setting a personalization indicator to on.

6. The method according to claim 1, wherein personalizing the mobile communication device comprises:

reading a first subscriber identity from a subscriber identity elementary file;

updating the subscriber identity elementary file with a second subscriber identity; and

reading the second subscriber identity from the subscriber identity elementary file.

7. The method according to claim 1, wherein personalizing the mobile communication device comprises:

issuing a single command to retrieve all subscriber identities from the single mobile communication device subscriber module;

storing the plurality of subscriber identities from the single mobile communication device subscriber module to a memory of the mobile communication device; and

setting a personalization indicator to on.

8. The method according to claim 1, wherein the plurality of subscriber identities are stored in a single elementary file on the single mobile communication device subscriber module.

9. The method according to claim 1, wherein the single mobile communication device subscriber module is a Subscriber Identity Module and the subscriber identity is an International Mobile Subscriber Identity.

10. A method in a mobile communication device including a single subscriber module having a plurality of subscriber identities, the method comprising:

issuing a select command to the single subscriber module, the select command selecting a subscriber identity elementary file on the single subscriber module, the subscriber identity elementary file containing the plurality of subscriber identities;

sending a read command to the single subscriber module;

receiving the plurality of subscriber identities from the single subscriber module in response to sending the read command;

storing the plurality of subscriber identities in a memory of the mobile communication device; and

setting a personalization indicator to on.

11. The method according to claim 10, further comprising:

operating the mobile communication device on a first network using a first subscriber identity;

detecting a change of network coverage to a second network; and

switching an operational subscriber identity from the first subscriber identity to a second subscriber identity based on the change of network coverage.

12. The method according to claim 10, further comprising receiving a response from the subscriber module in response to issuing the select command, the response including a file size of the subscriber identity elementary file

13. The method according to claim 10, wherein the read command includes an offset parameter indicating an offset in the subscriber identity elementary file, and a length parameter indicating a length of the data to be read.

14. The method according to claim 10, further comprising:

reading a subscriber identity from the single subscriber module;

comparing the subscriber identity with the plurality of subscriber identities stored in the mobile communication device; and

blocking use of selected features of the mobile communication device if the subscriber identity does not match one of the plurality of subscriber identities stored in the mobile communication device.

15. The method according to claim 10, wherein the subscriber module is a Subscriber Identity Module and the subscriber identity is an International Mobile Subscriber Identity.

16. A mobile communication device subscriber module comprising:

a controller configured to control the operations of the mobile communication device subscriber module;

an input and output contact point coupled to the controller;

a supply voltage contact point coupled to the controller;

a memory including a multiple subscriber identity elementary file, the multiple subscriber identity elementary file comprising a body including a plurality of subscriber identity locations and a plurality of subscriber identities, each of the plurality of subscriber identity locations comprising at least a subscriber identity of the plurality of subscriber identities, each subscriber identity consisting of eight bytes.

17 The mobile communication device subscriber module according to claim 16, wherein the multiple subscriber identity elementary file comprises a mandatory first subscriber identity of eight bytes.

18. The mobile communication device subscriber module according to claim 16, wherein the mobile communication device subscriber module is a Subscriber Identity Module and the subscriber identity is an International Mobile Subscriber Identity.

19. The mobile communication device subscriber module according to claim 16, wherein the controller is further configured to

operate the mobile communication device subscriber module on a first network using a first subscriber identity;

detect a change of network coverage to a second network; and

switch an operational subscriber identity from the first subscriber identity to a second subscriber identity based on the change of network coverage.

20. The mobile communication device subscriber module according to claim 16, wherein the controller is further configured to personalize a mobile communication device to the plurality of subscriber identities.

21. The mobile communication device subscriber module according to claim 16, wherein the controller is further configured to

receive a select command from a mobile communication device, the select command selecting the multiple subscriber identity elementary file;

accept a read command from the mobile communication device; and

send the plurality of subscriber identities from the subscriber module in response to accepting the read command.

22. The mobile communication device subscriber module according to claim 16, wherein the memory further includes a single subscriber identity elementary file comprising a body including a single subscriber identity.

23. A method in a mobile communication device including a plurality of subscriber identities on a single mobile communication device subscriber module, comprising:

storing the plurality of subscriber identities on the single mobile communication device subscriber module;

issuing a retrieve command for retrieving all of the plurality of subscriber identities on the single mobile communication subscriber module;

receiving all of the plurality of subscriber identities from the single mobile communication subscriber module in response to sending the read command; and

storing all of the plurality of subscriber identities to a memory of the mobile communication device.

24. The method according to claim 23, further comprising receiving a subscriber identity amount indicator, the subscriber identity amount indicator indicating a number of subscriber identities located on the single mobile communication subscriber module.

25. The method according to claim 23, wherein the subscriber module is a Subscriber Identity Module and the subscriber identity is an International Mobile Subscriber Identity.

26. The method according to claim 23, further comprising switching an operational subscriber identity from a first subscriber identity to a second subscriber identity.

27. The method according to claim 23, further comprising:

operating the mobile communication device on a first network using a first subscriber identity;

detecting a change of network coverage to a second network; and

switching an operational subscriber identity from the first subscriber identity to a second subscriber identity based on the change of network coverage.

28. The method according to claim 23, further comprising personalizing the mobile communication device to each of the plurality of subscriber identities by setting a personalization indicator to on.

29. The method according to claim 23, wherein the plurality of subscriber identities are stored in a single elementary file on the single mobile communication device subscriber module.

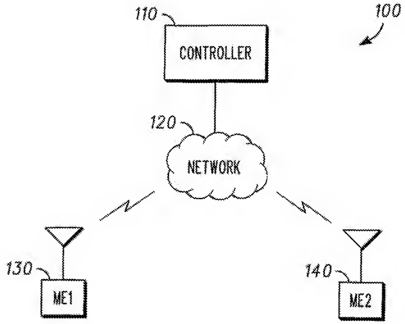

110
100
CONTROLLER
120
NETWORK
130
ME1
140
ME2

FIG. 1

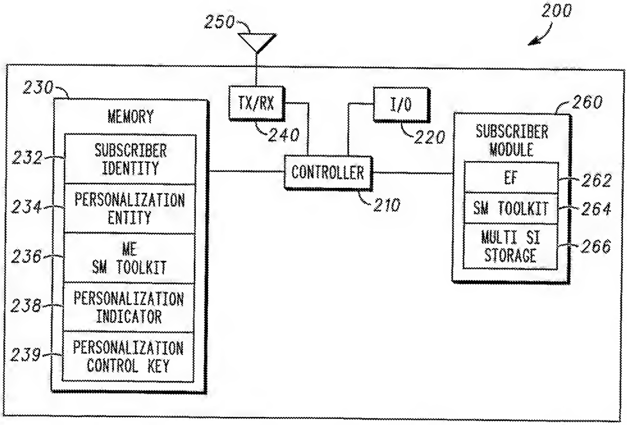

200
250
230
MEMORY
232
SUBSCRIBER IDENTITY
234
PERSONALIZATION ENTITY
236
ME SM TOOLKIT
238
PERSONALIZATION INDICATOR
239
PERSONALIZATION CONTROL KEY
TX/RX
240
I/O
220
CONTROLLER
210
260
SUBSCRIBER MODULE
EF
262
SM TOOLKIT
264
MULTI SI STORAGE
266

FIG. 2

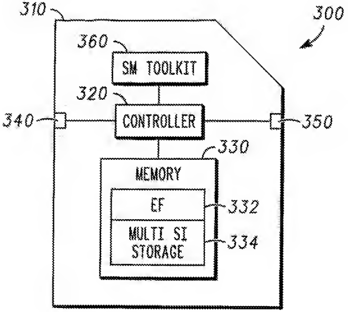

300
310
360
SM TOOLKIT
320
CONTROLLER
340
350
330
MEMORY
EF
332
MULTI SI STORAGE
334

FIG. 3

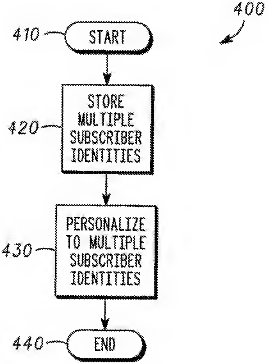

400
410
START
420
STORE MULTIPLE SUBSCRIBER IDENTITIES
430
PERSONALIZE TO MULTIPLE SUBSCRIBER IDENTITIES
440
END

FIG. 4

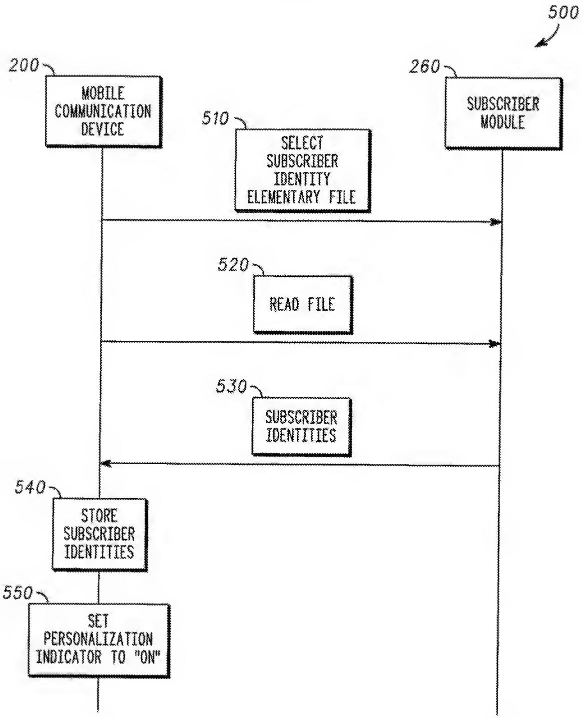

500
200
MOBILE COMMUNICATION DEVICE
260
SUBSCRIBER MODULE
510
SELECT SUBSCRIBER IDENTITY ELEMENTARY FILE
520
READ FILE
530
SUBSCRIBER IDENTITIES
540
STORE SUBSCRIBER IDENTITIES
550
SET PERSONALIZATION INDICATOR TO "ON"

FIG. 5

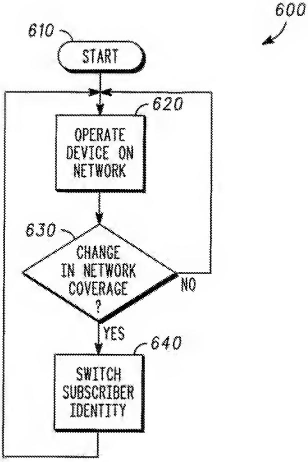

600
610
START
620
OPERATE DEVICE ON NETWORK
630
CHANGE IN NETWORK COVERAGE ?
NO
YES
640
SWITCH SUBSCRIBER IDENTITY

FIG. 6

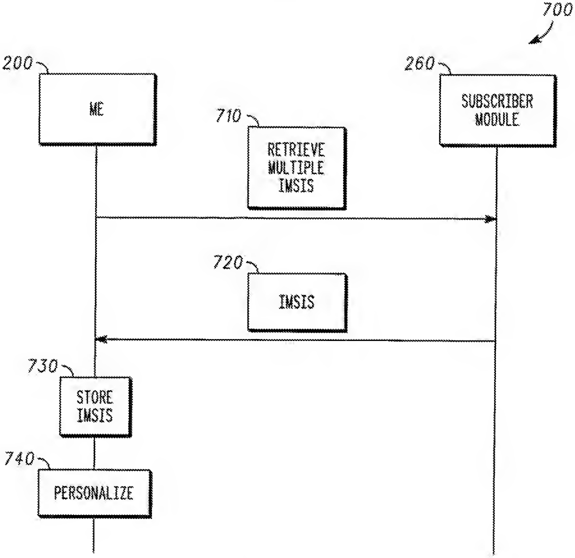

700
200
ME
260
SUBSCRIBER MODULE
710
RETRIEVE MULTIPLE IMSIS
720
IMSIS
730
STORE IMSIS
740
PERSONALIZE

FIG. 7